OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: December 28, 2002, 07:13:24 ; Search time 3144 Seconds
                   (without alignments)
                   8034.744 Million cell updates/sec

Title: US-09-921-994-1
Perfect score: 868
Sequence: 1 ataagaatgcgcgcccgcatga...........ctgggcaagtagaactagtcc 868

Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455402878 residues

Total number of hits satisfying chosen parameters: 4109280

Post-processing: Minimum Match 0%
Maximum DB seq length: 2000000000
Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries .

Database :   GenEmbl:*
1:    gb_ba:*
2:    gb_htg:*
3:    gb_in:*
4:    gb_om:*
5:    gb_ov:*
6:    gb_pat:*
7:    gb_ph:*
8:    gb_pl:*
9:    gb_pr:*
10:   gb_ro:*
11:   gb_sts:*
12:   gb_sy:*
13:   gb_un:*
14:   gb_vi:*
15:   em_ba:*
16:   em_fun:*
17:   em_hum:*
18:   em_in:*
19:   em_mu:*
20:   em_om:*
21:   em_or:*
22:   em_ov:*
23:   em_pat:*
24:   em_ph:*
25:   em_pl:*
26:   em_ro:*
27:   em_sts:*
28:   em_un:*
29:   em_vi:*
30:   em_htg_hum:*
31:   em_htg_inv:*
32:   em_htg_mus:*
33:   em_htg_other:*
34:   em_htg_pln:*
35:   em_htg_rod:*
36:   em_htg_mam:*
37:   em_htg_vrt:*
38:   em_sy:*
39:   em_htgo_hum:*
40:   em_htgo_mus:*
41:   em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query % Match	Length	DB	ID	Description
1		868	100.0	868	6	AX377958	AX377958 Sequence
2		686.2	79.1	1161	9	AR050784	AR050784 Sequence
3		686.2	79.1	1161	9	BC015364	BC015364 Homo sapi
4		683	78.7	1195	9	AR003351	AR003351 Sequence
5		683	78.7	1161	6	HUMEB13X	L08187 Human cytok
6		354	40.8	1131	10	BC008209	BC008209 Mus muscu
7		354	40.8	1161	10	AF013114	AF013114 Mus muscu
8		205.2	23.6	445	9	AX392987	AX392987 Sequence
9	c	182.6	21.0	43297	9	AC005578	AC005578 Homo sapi
10	c	172.2	19.8	179607	2	AC027272	AC027272 Homo sapi
11	c	172.2	19.8	43297	9	AC005578	
12	c	102.8	11.8	116	11	G54090	G54090 902 Human H
13	c	97	11.2	193355	2	AC073737	AC073737 Mus muscu
14	c	90.2	10.4	176359	2	AC118074	AC118074 Rattus no
15		74.8	8.6	1332	6	AX399177	AX399177 Sequence
16		74.8	8.6	1332	6	S54212	S54212 ciliary neu
17		73.2	8.4	125020	10	AF429315	AF429315 Homo sapi
18		69	7.9	1975	9	AF068615	AF068615 Mus muscu
19		69	7.9	1566	9	AX399178	AX399178 Sequence
20		69	7.9	1591	6	AR066232	AR066232 Sequence
21		69	7.9	1591	6	AR070290	AR070290 Sequence
22		69	7.9	1591	6	AR074863	AR074863 Sequence
23		69	7.9	10732	6	I12552	I12552 Sequence 1
24		69	7.9	1591	6	I56052	I56052 Sequence 1
25		69	7.9	1692	6	AX205060	AX205060 Sequence
26		69	7.9	1923	6	BC001492	BC001492 Homo sapi
27		69	7.9	2020	6	AX281744	AX281744 Sequence
28		67.4	7.8	1089	5	GGU29245	U29245 Gallus gall
29		65.8	7.6	1248	5	GGGPARA	248168 G.gallus mR
30	c	54.8	6.3	125020	9	AF429315	AF429315 Homo sapi
31		54.2	6.3	109523	2	AC111892	AC111892 Rattus no
32		53.8	6.2	114062	2	E32986	E32986 Gene encod1
33	c	52	6.0	3060	1	OSJN002267	AL731619 Oryza sat
34	c	51	5.9	159851	2	SAUREVTRAN	SAUREVTRAN Stigmatella
35	c	50.8	5.9	209887	2	AP003571	AP003571 Oryza sat
36	c	50.8	5.9	298166	2	AC079424	AC079424 Mus muscu
37	c	50.6	5.8	61958	2	AC087563	AC087563 Homo sapi
38	c	50.4	5.8	249262	2	AC121467	AC121467 Rattus no
39	c	49.8	5.7	303091	2	AC079430	AC079430 Mus muscu
40	c	49.8	5.7	303091	2	AC084799	AC084799 Mus muscu
41	c	49	5.6	7218	6	I66494	I66494 Sequence 14
42	c	49	5.6	158880	2	AC128264	AC128264 Rattus no
43	c	49	5.6	306695	2	AC079431	AC079431 Mus muscu
44	c	48.8	5.6	110000	2	AC098456_1	Continuation (2 of
45	c	48.4	5.6	130540	2	AC079417	AC079417 Mus muscu

ALIGNMENTS

RESULT 1
AX377958
LOCUS   AX377958   868 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION   Sequence 1 from Patent WO0212282.
ACCESSION   AX377958
VERSION   AX377958.1 GI:19574012
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bowman,M.R.
  TITLE   Novel ebi-3-alt protein and nucleic acid molecules and uses
               therefor

RESULT 2

```
LOCUS       AR050784                    1161 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5830451.
ACCESSION   AR050784
VERSION     AR050784.1  GI:5974148
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Devergne,O. and Kieff,E.D.
  TITLE     Haematopoietic cytokine Epstein Barr virus-induced protein
  JOURNAL   Patent: US 5830451-A 1 03-NOV-1998;
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="unknown"
ORIGIN
BASE COUNT      239 a    378 c    304 g    240 t
```

Query Match   79.1%; Score 686.2; DB 6; Length 1161;
Best Local Similarity   99.6%; Pred. No. 2.1e-128;
Matches 686; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

```
Qy        170 CATGAGCCCGCAGCTTCTCCTGGGCCCTTGTCCTCTGGCGCTGCCCCCCGGTGG 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 CATGAGCCCGCAGCTTCTCCTGGGCCCTTGTCCTCTGGCGCTGCCCCCCGGTGG 72

Qy        230 AAGGAAAAGGGCCCCCAGCAGCTCTGACAACTGCCCCGGGGTGCAATGCCCGGTA 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 AAGGAAAAGGGCCCCCAGCAGCTCTGACAACTGCCCCGGGGTGCAATGCCCGGTA 132

Qy        290 CCCGATCGCGGGAATTGCTCCTGGACCCCGTCCCAAAACTCCACCAGCCCCGT 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 CCCGATCGCGGGAATTGCTCCTGGACCCCGTCCCAAAACTCCACCAGCCCCGT 192

Qy        350 GTCCTTCATTGCCACGTACAGGCTCGGCAATGGGCCCCGGGCCACAGCTGGCCT 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 GTCCTTCATTGCCACGTACAGGCTCGGCAATGGGCCCCGGGCCACAGCTGGCCT 252

Qy        410 GCACGAGAGCCAACGTCCACCAGCTGCACCATCACCGCGGATGGTCAGTTTCCCATGGC 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 GCACGAGAGCCAACGTCCACCAGCTGCACCATCACCGCGGATGGTCAGTTTCCCATGGC 312

Qy        470 TCCCTACGTGCTCAAATGTCACGCGCGGTCCCGCCGGTGGGGCTCCGGCGTT 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 TCCCTACGTGCTCAAATGTCACGCGCGGTCCCGCCGGTGGGGCTCCGGCGTT 372

Qy        530 TTTCATTAACAGAGACAGCACATCATCAAGCCCGGACCCTCCAGAAAGGGCCCCT 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 TTTCATTAACAGAGACAGCACATCATCAAGCCCGGACCCTCCAGAAAGGGCCCCT 432

Qy        590 CGCTGAGCCCGCCAGCTCACAGTGAGTGGGAGGCTCTCTGGCCCTGGCCCTGCCT 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 CGCTGAGCCGCCAGCTCACAGTGAGTGGGAGGCTCTCTGGCCCTGGCCCTGCCT 492

Qy        650 CTTCTCACTGAAGTACTGGATCCGTTACAAGGCGTCAGGGAGCGTTCGTGCC 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 CTTCTCACTGAAGTACTGGATCCGTTACAAGGCGTCAGGGAGCGTTCGTGCC 552

Qy        710 GGGGGCCCATTGAAGCCACGTCCTTCATCCTCAGGGGCTGTGTGGCTTCGTACTA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 GGGGGCCCATTGAAGCCACGTCCTTCATCCTCAGGGGCTGTGTGGCTTCGTACTA 612

Qy        770 CGTCCAAGTGGGCGGCTCAGGAGCTCACAGACTACGGGGAAACTGAGTGAGTCTCC 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 CGTCCAAGTGGGCGGCTCAGGAGCTCACAGACTACGGGGAAACTGAGTGAGTCTCC 672

Qy        830 CGCCACTGCCACAATGAGCCCTGGGCAAGTAG 860
              |||||||||||||||||||||||||||||||
Db        673 CGCCACTGCCACAATGAGCCCTGGGCAAGTAG 703
```

```
JOURNAL     Patent: WO 0212282-A 1 14-FEB-2002;
            GENETICS INSTITUTE, INC. (US)
FEATURES             Location/Qualifiers
     source          1..868
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
ORIGIN
BASE COUNT      147 a    325 c    239 g    157 t
```

Query Match   100.0%; Score 868; DB 6; Length 868;
Best Local Similarity   100.0%; Pred. No. 5.1e-165;
Matches 868; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 ATAAGAATGCGGCGCCAGCCAGCCCCGGCCCTTGTCCTCTGGGGCCAGCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATAAGAATGCGGCGCCAGCCAGCCCCGGCCCTTGTCCTCTGGGGCCAGCTG 60

Qy         61 CCCGCCCCCAGAGTGGAAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCAATG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCGCCCCCAGAGTGGAAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCAATG 120

Qy        121 CCGAGGCCTTCGGGTACCCGATCGCCGTGGATTGCTCCTGGACCCCGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGAGGCCTTCGGGTACCCGATCGCCGTGGATTGCTCCTGGACCCCGC 180

Qy        181 AGCTTCTTCCTGGCCCTTGTCTCTGGGCCAGCTGCGCGTGAGGAAAGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCTTCTTCCTGGCCCTTGTCTCTGGGCCAGCTGCGCGTGAGGAAAGGGC 240

Qy        241 CCCCAGCAGCTCTGACAACTGCCCCGGGGTGCAATGCCCGGTACCCGATCGCGGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCCAGCAGCTCTGACAACTGCCCCGGGGTGCAATGCCCGGTACCCGATCGCGGG 300

Qy        301 TGGAATTGCTCCTGGACCCCGTCGCCGCTGCTCCAAAACTCCACCAGCCCCGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGAATTGCTCCTGGACCCCGTCGCCGCTGCTCCAAAACTCCACCAGCCCCGTG 360

Qy        361 CCACGTACAGGCTCGGCAATGCTGCCCGGGCCACAGCTGGCCTGCACGAGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACGTACAGGCTCGGCAATGCTGCCCGGGCCACAGCTGGCCTGCACGAGAGC 420

Qy        421 CAACGTCCACCAGCTGCACCATCACGGATGGTCCAGTGTTCTCCATGGGCTCCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAACGTCCACCAGCTGCACCATCACGGATGGTCCAGTGTTCTCCATGGGCTCCGT 480

Qy        481 TCAAATGTCACGCGCCGTCCCGCCGGGGTCCAGCAGCAGCTTCGTGCTTTCATTAAACAG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCAAATGTCACGCGCCGTCCCGCCGGGGTCCAGCAGCAGCTTCGTGCTTTCATTAAACAG 540

Qy        541 AGCACATCATCAAAGCCCGGACCCTCCAGAAAGGCGTGCGGCCTAAGCCCGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGCACATCATCAAAGCCCGGACCCTCCAGAAAGGCGTGCGGCCTAAGCCCGCC 600

Qy        601 AGCTACACAGGGTGCAGTGGGAGGACCTCCCGGGGTCTGGCCCTTCCCAGAATCTTCTCACTGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGCTACACAGGGTGCAGTGGGAGGACCTCCCGGGGTCTGGCCCTTCCCAGAATCTTCTCACTGA 660

Qy        661 AGTACTGGATCCGTTACAAGGCGTCAGGGAGCGTTCGTGCCCCCATTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGTACTGGATCCGTTACAAGGCGTCAGGGAGCGTTCGTGCCCCCATTG 720

Qy        721 AAGCCACGTCCTTCATCCTCAGGGGCTGTGTGCCGGGCGTGTGGCTTCATCCTCAGGGGTGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAGCCACGTCCTTCATCCTCAGGGGCTGTGTGCCGGGCGTGTGGCTTACTACGTCCAAGTGG 780

Qy        781 CGGCTCAGGGACCTCACAGACTACGGGGAAACTGAGTGAGTCACTGGAGTGACTAGTGCCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGGCTCAGGGACCTCACAGACTACGGGGAAACTGAGTGAGTCACTGGAGTGACTAGTGCCA 840

Qy        841 CAATGAGCCCTGGGCAAGTAGACTAGTCC 868
              |||||||||||||||||||||||||||||
Db        841 CAATGAGCCCTGGGCAAGTAGACTAGTCC 868
```

RESULT 3

```
LOCUS       BC015364
DEFINITION  Homo sapiens, Epstein-Barr virus induced gene 3, clone MGC:21515
            IMAGE:3897136, mRNA, complete cds.
ACCESSION   BC015364
VERSION     BC015364.1  GI:15929884
KEYWORDS    .
SOURCE      MGC.
  ORGANISM  Homo sapiens
            Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1195)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT     Contact: MGC Project URL: http://mgc.nci.nih.gov
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
REMARK      Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 22 Row: f Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1..1195
                     /organism="Homo sapiens"
                     /db_xref="LocusID:10148"
                     /db_xref="taxon:9606"
                     /clone="MGC:21515 IMAGE:3897136"
                     /tissue_type="Pancreas, epithelioid carcinoma"
                     /clone_lib="NIH_MGC_70"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             54..743
                     /codon_start=1
                     /product="Epstein-Barr virus induced gene 3"
                     /protein_id="AAH15364.1"
                     /db_xref="GI:15929885"
                     /translation="MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVGCRASRYPIAV
                     DCSWTLPPARNSTSPVSFIATYRLGMAARGHSWPCLQQTPTSTSCTITDVQLFSMAPY
                     VLNVTAVHPWGSSSSFVPFITEHIIKPDPEPGVRLSPLAERQLQVQWEPPGSWPFPEI
                     FSLKYWIRYKRQGAARFHRVGPIEATSFILRAVRPRARYVQVAAQDLITDYGELSDWS
                     LPATATMSLGK"
BASE COUNT      247 a      395 c      311 g      242 t
ORIGIN
```

Query Match 79.1%; Score 686.2; DB 9; Length 1195;		
Best Local Similarity 99.6%; Pred. No. 2e-128;		
Matches 688; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		

```
Qy    170 CATGACCCCGCCAGCGTTCTCCTGGCCCCTTGTCCCTTGGGCCCAGCTGCCCCCTGCGGTGG 229
          ||||||||||||||||||||||||||||||| |||||||||||||||| ||||||||| ||||
Db     53 CATGACCCCGCCAGCGTTCTCCTGGCCCCTTGTCCCTTGGGCCCAGCTGCCCCCTGCAGTGG 112

Qy    230 AAGGAAAGGGCCCCAGCAGCTCTGACACTGCCCCCGGGGTGCAATGCCGAGCCCTCTCGGTA 289
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 AAGGAAAGGGCCCCAGCAGCTCTGACACTGCCCCCGGGGTGCAATGCCGAGCCCTCTCGGTA 172
```

(right column)

```
Qy    290 CCCGGATCGCGCCGTGGGATTGCTCCCTGGACCCTGCCGCCGCGGCCTGCTCCAAAACTCCACCAGCCCCGT 349
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    173 CCCGGATCGCGCCGTGGGATTGCTCCCTGGACCCTGCCGCCGCGGCCTGCTCCAAAACTCCACCAGCCCCGT 232

Qy    350 GTCCTTCATTGCCACGTACAGAGCTCGGCATGGCTCCCCGGGGCCCACAGCTGGGCCTTGCCT 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 GTCCTTCATTGCCACGTACAGAGCTCGGCATGGCTCCCCGGGGCCCACAGCTGGGCCTTGCCT 292

Qy    410 GCAGCAGACCCCAAGGTCACCACCAGTGCACCATCACGGATGTCCAGCTGTTCTCCATGGCC 469
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 GCAGCAGACCCCAAGGTCACCACCAGTGCACCATCACGGATGTCCAGCTGTTCTCCATGGCC 352

Qy    470 TCCCTACGTGCCTCAAATGTCACCGCCGTGCACCCGTGGGGCTCCAGCAGCAGCTTCGTGCC 529
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 TCCCTACGTGCCTCAAATGTCACCGCCGTGCACCCGTGGGGCTCCAGCAGCAGCTTCGTGCC 412

Qy    530 TTTCATTAACAGAGCACATCATCAAGCCCGACCCTGCCCTCCAGAAGGGCCTAAGCCCCT 589
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 TTTCATTAACAGAGCACATCATCAAGCCCGACCCTGCCCTCCAGAAGGGCCTAAGCCCCT 472

Qy    590 CGCTCAGGCGCCCAGCGTACAGGTTGCAGTGGGAGAGCCTCCCCGGGGTCCTGGCCCTTTCCCAGAGAT 649
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 CGCTGAGGCGCCCAGCGTACAGGTTGCAGTGGGAGAGCCTCCCCGGGGTCCTGGCCCTTTCCCAGAGAT 532

Qy    650 CTTCTCCACTGAAGTACTGGATCCGGTTACAAGGCGTCAGGGAGCGTCGCGGCCTTCACCGGGGT 709
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 CTTCTCCACTGAAGTACTGGATCCGGTTACAAGGCGTCAGGGAGCGTCGCGGCCTTCACCGGGGT 592

Qy    710 GGGGCCCATTGAAGCCACGTCCTTCATCCTCCAGGGCGTCGTCCGGCCCCAGGTACTA 769
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    593 GGGGCCCATTGAAGCCACGTCCTTCATCCTCCAGGGCGTCGTCCGGCCCCAGGTACTA 652

Qy    770 CGTCCAAGTGGCGGCGCTCAGGACCCTCACACAGCTACGGGGAACTGAGTGACTGGAGTCTCCC 829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 CGTCCAAGTGGCGGCGCTCAGGACCCTCACAGACTACGGGGAACTGAGTGACTGGAGTCTCCC 712

Qy    830 CGCCACTGCCACAATGAGCCTGGGCAAGTAG 860
          ||||||||||||||||||||||||||||||||
Db    713 CGCCACTGCCACAATGAGCCTGGGCAAGTAG 743
```

RESULT 4

```
LOCUS       AR003351                1161 bp    DNA     linear   PAT 04-DEC-1998
DEFINITION  Sequence 5 from patent US 5744301.
ACCESSION   AR003351
VERSION     AR003351.1  GI:3964610
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Birkenbach,M. and Kieff,E.
  TITLE     Methods of detection of epstein barr virus induced genes expressed
            in the placenta
  JOURNAL   Patent: US 5744301-A 5 28-APR-1998;
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="unknown"
BASE COUNT      239 a      378 c      304 g      240 t
ORIGIN
```

Query Match 78.7%; Score 683; DB 6; Length 1161;		
Best Local Similarity 99.3%; Pred. No. 9e-128;		
Matches 686; Conservative 0; Mismatches 5; Indels 0; Gaps 0;		

```
Qy    170 CATGACCCCGCCAGCGTTCTCCTGGCCCCTTGTCCCTTGGGCCCAGCTGCCCCCTGCGGTGG 229
          ||||||||||||||||||||||||||||||| |||||||||||||||| ||||||||| ||||
Db     13 CATGACCCCGCCAGCGTTCTCCTGGCCCCTTGTCCCTTGGGCCCAGCTGCCCCCTGCAGTGG 72

Qy    230 AAGGAAAGGGCCCCAGCAGCTCTGACACTGCCCCCGGGGTGCAATGCCGAGCCCTCTCGGTA 289
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 AAGGAAAGGGCCCCAGCAGCTCTGACACTGCCCCCGGGGTGCAATGCCGAGCCCTCTCGGTA 289
```

```
Db   73 AAGGAAAAGGGCCCCCCAGCAGCTCTGACACTGCCCGGGGTGCAATGCCGAGCCTTCGGGTA 132
Qy  290 CCCGGATCGCCGTGGATTGCTCCTTGGACACTGCCC                          349
        |||||||||||||||||||||||||||||||||||
Db  133 CCCGGATCGCCGTGGATTGGCTCCTTGGA                                  192
Qy  350 GTCCTTCATTGCCACGTACAGGTCTTCGGG                                409
Db  193 GTCCTTCATTGCCACGTACAGGTCTTCGGG                                252
Qy  410 GCAGCAGCGACCCAACGTCCACCAGCTGCACC                              469
Db  253 GCAGCAGCGACCCAACGTCCACCAGCTGCACC                              312
Qy  470 TCCCTACGTGCTCAATGTCACCGCCGT                                   529
Db  313 TCCCTACGTGCTCAATGTCACCGCCGT                                   372
Qy  530 TTTCATAACAGAGCACATCATCAAGCCC                                  589
Db  373 TTTCATAACAGAGCACATCATCAAGCCC                                  432
Qy  590 CGCTGAGCGCCCAGCTACAGGTGCAGTGG                                 649
Db  433 CGCTGAGCGCCCACGTACAGGTGCAGTGG                                 492
Qy  650 CTTCTCACTGAAGTACTGGATCCGTTACA                                 709
Db  493 CTTCTCACTGAAGTACTGGATCCGTTACA                                 552
Qy  710 GGGGGCCCATTGAAGCCACGTCGTCCT                                   769
Db  553 GGGGGCCCATTGAAGCCACGTCGTCCT                                   612
Qy  770 CGTCCAAGTGGCGGCGTCAGGACCTCACA                                 829
Db  613 CGTCCAAGTGGCGGCGTCAGGACCTCACA                                 672
Qy  830 CGCCCACTGCCACAATGAGCCTGGGCAAGTAG 860
Db  673 CGCCACTGCCACAATGAGCCTGGGCAAGTAG 703
```

RESULT 5
HUMEBI3X
LOCUS   HUMEBI3X   1161 bp   mRNA   linear   PRI 01-APR-1996
DEFINITION Human cytokine receptor (EBI3) mRNA, complete cds.
ACCESSION   L08187
VERSION   L08187.1 GI:632973
KEYWORDS   cytokine receptor.
SOURCE   Homo sapiens cDNA to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1161)
AUTHORS   Devergne,O., Hummel,M., Koeppen,H., Le Beau,M.M., Nathanson,E.C.,
            Kieff,E. and Birkenbach,M.
TITLE   A novel interleukin-12 p40-related protein induced by latent
            Epstein-Barr virus infection in B lymphocytes
JOURNAL   J. Virol. 70 (2), 1143-1153 (1996)
MEDLINE   96135230
PUBMED   8551575
COMMENT   On Jan 24, 1995 this sequence version replaced gi:181921.
FEATURES   Location/Qualifiers
     source   1..1161
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="BL41;B95-8"
                     /cell_type="B lymphocyte, EBV-converted Burkitt lymphoma"
     gene   1..1161
                     /gene="EBI3"
     CDS   14..703
                     /gene="EBI3"
                     /codon_start=1
                     /product="cytokine receptor"
                     /protein_id="AAA93193.1"
                     /db_xref="GI:632974"
                     /translation="MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAV
                     DCSWTLPPAPNSTSPVSFIATYRLGMAARGHSWPQCLQQTFPTSTSCTITDVQLFSMAPY
                     VLNVTAVHPWGCSSSSPVPFITEHIIKPDPPEGVRLSPLAERHVQVQWEPPGSWPFPEI
                     FSLKYWIRYKRQGAARFHRVGPIEATSFILRAVRPRARYVVQVAAQDLTDYGELSDWS"
     sig_peptide   14..73
                     /gene="EBI3"
     mat_peptide   74..700
                     /gene="EBI3"
                     /product="cytokine receptor"
BASE COUNT   239 a   378 c   304 g   240 t
ORIGIN

Query Match 78.7%; Score 683; DB 9; Length 1161;
Best Local Similarity 99.3%;
Matches 686; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy  170 CATGAACCCGCAGCCTTCTCCCTGCCCCTTGTCCCTGGGCCAGCTGCCGGTGG 229
        ||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db   13 CATGAACCCGCAGCCTTCTCCCTGCCCCTTGTCCCTGCCCAGCTGCCAGTGG  72
Qy  230 AAGGAAAAGGGCCCCCCAGCAGCTCTGACACTGCCC              289
        ||||||||||||||||||||||||||||||||||||
Db   73 AAGGAAAAGGGCCCCCCAGCAGCTCTGACACTG                 132
Qy  290 CCCGGATCGCCGTGGATTGCTCCTTGGACACTGCCC              349
        ||||||||||||||||| ||||||||||||||||||
Db  133 CCCGGATCGCCGTGGATTGGCTCCTTGGA                     192
Qy  350 GTCCTTCATTGCCACGTACAGGTCTTCGGG                    409
        |||||||||||||||||||||||||||||
Db  193 GTCCTTCATTGCCACGTACAGGTCTTCGGG                    252
Qy  410 GCAGCAGCGACCCAACGTCCACCAGCTGCACC                  469
        ||||||||||||||||||||||||||||||||
Db  253 GCAGCAGCGACCCAACGTCCACCAGCTGCACC                  312
Qy  470 TCCCTACGTGCTCAATGTCACCGCCGT                       529
        |||||||||||||||||||||||||||
Db  313 TCCCTACGTGCTCAATGTCACCGCCGT                       372
Qy  530 TTTCATAACAGAGCACATCATCAAGCCC                      589
        ||||||||||||||||||||||||||||
Db  373 TTTCATAACAGAGCACATCATCAAGCCC                      432
Qy  590 CGCTGAGCGCCCAGCTACAGGTGCAGTGG                     649
        ||||||||||||| ||||||||||||||
Db  433 CGCTGAGCGCCCACGTACAGGTGCAGTGG                     492
Qy  650 CTTCTCACTGAAGTACTGGATCCGTTACA                     709
        |||||||||||||||||||||||||||||
Db  493 CTTCTCACTGAAGTACTGGATCCGTTACA                     552
Qy  710 GGGGGCCCATTGAAGCCACGTCGTCCT                       769
        ||||||||||||||||||||||||||
Db  553 GGGGGCCCATTGAAGCCACGTCGTCCT                       612
Qy  770 CGTCCAAGTGGCGGCGTCAGGACCTCACA                     829
        |||||||||||||||||||||||||||||
Db  613 CGTCCAAGTGGCGGCGTCAGGACCTCACA                     672
Qy  830 CGCCCACTGCCACAATGAGCCTGGGCAAGTAG 860
        |||| ||||||||||||||||||||||||||
Db  673 CGCCACTGCCACAATGAGCCTGGGCAAGTAG 703
```

RESULT 6
BC008209
LOCUS   BC008209   1131 bp   mRNA   linear   ROD 07-AUG-2002

```
DEFINITION   Mus musculus, Epstein-Barr virus induced gene 3, clone MGC:11457
             IMAGE:3154957, mRNA, complete cds.
ACCESSION    BC008209
VERSION      BC008209.1  GI:14198294
KEYWORDS     MGC.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1131)
  AUTHORS    Strausberg,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (22-MAY-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Baylor College of Medicine Human Genome
             Sequencing Center
             Center code: BCM-HGSC
             Web site: http://www.hgsc.bcm.tmc.edu/cdna/
             Contact: amg@bcm.tmc.edu
             Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
             Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
             Richards, S., Gibbs, R.A.
             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 16 Row: c Column: 19
             This clone was selected for full length sequencing because it
             passed the following selection criteria: matched mRNA gi: 7657051.
FEATURES             Location/Qualifiers
     source          1..1131
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="129,C57BL/6J,FVB/N"
                     /clone="MGC:11457 IMAGE:3154957"
                     /tissue_type=Mammary tumor. Brca1-/fl; MMTV-Cre model. 10
                     months old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam3"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             188..874
                     /codon_start=1
                     /product="Epstein-Barr virus induced gene 3"
                     /protein_id="AAH08209.1"
                     /db_xref="GI:14198295"
                     /db_xref="LocusID:50498"
                     /translation="MSKLLFLSLALWASRSPGYTETALVALSQPRVQCHASRYPVAVD
                     CSWTPLQAPNSTRSTSFIATYRLGVATQQQSQPCLQRSPQASRCTIPDVHLFSTVPYM
                     LNVTAVHPGCASSLLAFVAERIIKPDPEGVRLRTAGQRLQVLWHPPASWPPDIFS
                     LKYRLRYRRGASHFRQVGPIEATTFLRNSKPHAKYCIQVSAQDLTDYGKPSDWSLP
                     GQVESAPHKP"
BASE COUNT       242 a       374 c       268 g       247 t
ORIGIN
```

```
Qy          22 CCCCCAGCGTTCTCCTGGCCCCCCGTCGTCCTCTGGGCCGGGCCCCGCCCCCCCCCGCCAGTGGAAGGAA 81
               |||| || ||||| || ||  ||| | |||| |||| ||  ||  |||| |||||||||||||||||
Db          39 CCATCATCACTCTCGCCCCATACACTGACCACACTGAGCCACACTGGGCGCAGGTCCTCTTCCC 98

Qy          82 AGGGCCCCCAGCAGCTCTGACACTGCCCCGGGTGCCAATGCCGAGCCTCTCGGGTACCCAGT 141
               ||  |||| ||  ||  ||||   |||| || | ||  | ||  |||||| || ||| |||
Db          99 TGGGGCCCAGGTTCCTGTCTGTGTGAGTCCCCTCTGATGGGTCACTAA 158
```

```
RESULT 7
AF013114
LOCUS       AF013114     1161 bp   mRNA   linear   ROD 21-AUG-1997
DEFINITION  Mus musculus cytokine receptor-like molecule (EBI3) mRNA, complete
            cds.
ACCESSION   AF013114
VERSION     AF013114.1  GI:2338440
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Nomura,H., Yaguchi,N., Hanyuu,C., Maeda,M., Kikuchi,Y., Nakata,Y.,
            Kojima,T., Tulin,E.E. and Hasegawa,M.
  TITLE     Mouse homolog of human Epstein-Barr virus-induced gene 3 (EBI3)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1161)
  AUTHORS   Nomura,H., Yaguchi,N., Hanyuu,C., Maeda,M., Kikuchi,Y., Nakata,Y.,
            Kojima,T., Tulin,E.E. and Hasegawa,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1997) Cytokine Research Program, CHUGAI Research
            Institute for Molecular Medicine, Inc., 153-2, Nagai, Niihari,
            Ibaraki 300-41, Japan
```

```
Qy         142 CGCCCTTGGATTGCTCCTGGACCCCTGTGGCCATGACCCCCAGCGTTCTCCTGGCCCTTGTCC 201
               ||| | ||| || ||| ||| ||||||  ||  ||||||||||||||||||||||||| |||
Db           1 CTCCGGATCCAAGGAACAGAGACCACAGAGCATGTCCAAGCTGCTTTCCTTGTCACTTGCC 218

Qy         202 TCTGGGCCAGCTGCCCGGCCCCCGGCCCCCCCGCGGTGGAAGGAAAGGGCCCCGGCCATGG 261
               |||||||||||||||| |  |||||| | |||||     || | |||||||| | ||
Db         219 TCTGGGCACAGCCGGCCCCCT---GGTTACACTGAAACAGCTCGTTCGTGGCTCTAAGCCAGC 275

Qy         262 CCCGGGTGCAATGCCGAGCCTCTCGGGTACCCAGTGGGATTGCGTTGCTGCTCCCTTGC 321
               | |||||| || | || || || || ||||| ||  ||| ||| || | || || |||
Db         276 CCAGAGTGCAAATGCCATGGCTTCTCGGTATCCCGTGGCCGTGGACTGCTCCTGGACTCC 335

Qy         322 CGCCTGCTCCAAAACTCACCACCAGCCCCGGTGTCCTTCATTGCCACGTACAGGCTCGGGATGG 381
               | || | | |   |  ||||| || || ||||  | |||  ||||| |||  || | |||
Db         336 TCCAGGCTCCAACTCCACCAGATCCACCTCTTCATTGCCACTTACAGGCTCGGTGTGG 395

Qy         382 CTGCCCGGGGGTCCAGCACAGCTTGCCTGTCTTTCATCAACAGACACAATCATCAACAGCCCCAGACC 441
               || |||| || ||||||| | | || ||| ||| || |  || | |  | || | || ||||
Db         396 CCACCCAGCAGCAGCCAGCCCTGCCTACAACAGGAGCCCCCAGGCCTCCCGAGCCAGCACCA 455

Qy         442 TCACGGATGGTCCAGCTGTTCTCCACGGTTCCATGGCTCCCTACGTGCTCAATGTCACCGCCGGCGCC 501
               |||| || ||||| |||  | ||| | || || |||||   | ||  |    || |  | |||
Db         456 TCCCCGACGTGCACCTGTTCTCACGGTCCCCTACATGCTAAATGTCACTGCAGTTCACC 515

Qy         502 CCTGGGGCCTCCAGCAGCCACTTGTCGCTTTGTCATTCATCAACAGACACAATCATCAACAGCCCCAGACC 561
               |||| || ||||||| |  ||||| || ||| |   ||| || |  ||  || | | || ||||
Db         516 CAGGCGCGGGCCAGCAGCAGCAGCAGCGCTCCTTAGCCTTTGTTGGCTGAGAATCATCAAGCCAGGACC 575

Qy         562 CTCCAGAAGCAGCGGTCGGCTGCGGACGCCCCTCGGCTGAGCGCCGGCTACAGGTGGGGAGC 621
               ||  ||  || |  | || || ||  |  || |  | || || |  || |  ||| |||  ||
Db         576 CTCCGGAAGCGGTCGCCTGGCCACAGCGGGGACAGCG---CTGCCAGGTGCTCTGGCATC 632

Qy         622 CTCCCGGGTGTCCTGGCGGCCCCTTCCCAGAGATCTTCTCACTGAAGTACTGGATCCGTTACAAGC 681
               || ||| | || |||| |    |||||||| || || || | |   || ||| |  || ||| ||
Db         633 CCCCTGCTTCCTTGGCCCTCTCTCCGGACACTTTCTCTCAAGTACCGACTCCGGTACCGGC 692

Qy         682 GTCAGGGAGCTGCGCGGTCTTCCACGGGGTGGGGCCCATTGAAGCCACGTCCTTCATCCTCA 741
               ||| | || | | || | |  |||| ||| ||||  || ||| | || | |  ||  | |
Db         693 GCCGAGGAGGCTCACTTCCGCCAGGTGGGACCCATTGAAGCCACGACTTTCACCCTCA 752

Qy         742 GGGCTGGTGGGCGGCCCCGGAGCCCAGGTGACTACGTCCAAGTGGGCCTCAGGAGCCTCACAGACT 801
               || ||  | || || |  ||| |||||| | ||| | |||||  || |||| ||| |  | |||
Db         753 GGAACTCGAAAACCCCATGCCAAGTATTGCATCCAGGTGTCAGCTCAGGACCTCACAGAGATT 812

Qy         802 ACGGGAAGCTGAGTGACTGGAGTCTCCCCG 831
               | ||||||   ||  |||||| ||| ||| |
Db         813 ATGGGAAACCAAGTGACTGGAGCCTCCCTG 842
```

```
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="Clontech Marathon-Ready library #7459-1"
                     /dev_stage="15-day embryo"
     gene            1..1161
                     /gene="EBI3"
                     /note="MMEBI3"
     CDS             223..909
                     /gene="EBI3; MUSEBI3"
                     /note="similar to human EBI3 protein"
                     /codon_start=1
                     /product="cytokine receptor-like molecule"
                     /protein_id="AAB67115.1"
                     /db_xref="GI:2338441"
                     /translation="MSKLLFLSLALWASRSPGYTETALVALSQPRVQCHASRYPVAVD
                     CSWTLPQAPNSTRSTSFIATYRLGVATQQOSQPCLQRSPQASRCTIPDVHLFSTVPYM
                     LNTAVHPGGASSSLLAFVWAERIIKPDPPEGVRLRTAGQRLQVLWHPPASWPFPDIFS
                     LKYRLRYRRGASHFRQVGPIEAFTFLRNSKPHAKYCIQVSAQDLIDYGKPSDWSLP
                     GQVESAPHKP"
BASE COUNT      244 a     390 c     272 g     255 t
ORIGIN
```

```
Query Match     40.8%;  Score 354;  DB 10;  Length 1161;
Best Local Similarity   66.5%;  Pred. No. 1.7e-61;
Matches 539;  Conservative   0;  Mismatches 265;  Indels   6;  Gaps   2;
```

```
Qy     22 CCCGCGAGCTTCTCCTGCGGCCCCTTGTCCTCTGGGCCAGCTGCCCCGGCCCCCCCGGCGTGAAGGAA 81
          |||| |||  |||||        |||||||| ||||||||||||||    |||||||  |||||| |||
Db     74 CCATCACATCTCCCGCCCCATACACTGACACTGAGCGACCACACTGGGGCAGGTGCCTTCCC 133

Qy     82 AGGGGCCCCCAGCAGCTCTGACACTGGGTCCCCTGTGTGGTGAAGTCCTCGGGTTCACTAA 141
          ||||  |  ||||| |||||  |||||  |||||||||||    |||||    ||||| ||
Db    134 TGGGGCCAGGTTCCCTGTGTGAGTCCGTCTTCACCCTCTCTGAGTGGGTCACTA 193

Qy    142 CGCCGTGGAATTGCTTCCTGGACCCTGCCCAGTGACCCCGAGCTTCTCTGGCCTTTGTCC 201
          || | |||||||  |  |||   |||  ||||||| |||||||||||| |||  |||||
Db    194 CTCGGAATCCAAGGAACAGAGGCCACAGAGCATGTCCAAGCTGCTTCCTGTCACTTGCC 253

Qy    202 TCTGGGCCAGCTGCCCCGCCCCGGTGGAGGAAAAGGGCCCCAGCTCTGACACTGC 261
          || |||||||||| ||| || ||| ||||| |||  |||||||||||  |||| |
Db    254 TCTGGGGCCAGCCGCCTCCCT---GGTTACACTGAAACAGCTTCTCGTGGCCTCTAAGCCAGC 310

Qy    262 CCCGGGGCTGCAAGTGCCGAGCCTCTCGGCTACCGGACGCGCCGTGGATTGCTCCTGGACCCTGC 321
          ||| ||| ||||| | ||  ||||| ||| |||  || | |||  || ||||||||||||| ||
Db    311 CCAGAGTGGCAANGCCANGCCTTCTCGGTATCCCGTGCCGTGGACTGCCTCTGGGACTCCTC 370

Qy    322 CGCCTGCTCCAAAACTCCACAGCACCGCGTCGTTCATTGCCAGGCTGTACAGGCTGGCATGG 381
          ||  ||||| |||  ||||||||| |||| | |||  |||| || ||  |||| |||  |||
Db    371 TCCAGGCTCCAACTCCACACGCCCGTGCCCACTTTCACTTACAGCTCGGTTCG 430

Qy    382 CTTGCCCGGGGCACACGCTGGCCTTGCCTGCAGCAGCAGCGTCCACAGCTGCCACCA 441
          |||| |||||| |||| | |||| |||| |||  ||||  | |||||||| ||| ||
Db    431 CCACCCAGCAGCAGGCCCTGCCTACAAACGGAGCCCCCGGATGCCACCA 490

Qy    442 TCACGGATGTGCCAGCTGTTCTCCATGGCTCTCCCTACGGTGCTCAAATGTCACGCGCCGTTCCGCC 501
          |||  |  ||  ||||||||| |||||||| ||||  | | ||  |||  | || |  |||||| |
Db    491 TCCCCGACGTGCACCTGTTCTCCACGGTGCCCTACATGCTAAATGTCACTGCAGTGCACC 550

Qy    502 CCTGGGGCCTCCAGCACGAGCTTGCGTGCCTTTGCATAAACAGAGCACATCATCAAGCCCGACC 561
          ||||||  ||||||||| ||| |||  |||  |  |||||  || |  ||| |||||||||||| |
Db    551 CAGGGGCGCGCCAGCGACCTCCTAGCCTTTGTGTGGGCAGAGCGCATCATCAAGCCCGACC 610

Qy    562 CTCCAGAAGGCGTGCGCCTAAGCACCGCCGGAATCATCAAGCCGGAGC 621
          ||||  ||||| | ||||| |||| || ||||  || |||||| |||
Db    611 CTCCGGAAGGCGTGCGCCTGCGCACAGCCGGACAGCGC---CTGCAGGTCCTCTGGCATC 667

Qy    622 CTCCGGGGCTCCTGGCCCTTGGCCCTTCTCCCAGAGATCTTCTCACTGAAGTACTGGGATCGGTTACAGC 681
          || || ||||||||||| ||||  ||||||||||| ||| ||||||||| ||| |||| |  |||| |||
Db    668 CCCCTGCTTCTCTGGCCCTTCTCCCGGACAATCTTCTCAAGTACCACCGACTCCGGCTACCGGC 727

Qy    682 GTCAGGGAGCGCTGCGCGCCTTCACGACGTGGGCCCCATTGAAGCCAGTGTCCTTCATCCTCA 741
          || |||  | |||  |||||||||| |||  ||| || |  ||||| ||  ||||||||||
Db    728 GCCAGGAGGAGCCCTCACTTCCGCGACTTCCGGCAGGTGGCCAGTGTCGGCGTCTGGCATCTGGACCCTCA 787
```

```
RESULT 8
AX392987
LOCUS       AX392987              445 bp    DNA        linear    PAT  23-MAR-2002
DEFINITION  Sequence 29 from Patent WO0214358.
ACCESSION   AX392987
VERSION     AX392987.1   GI:19701034
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Edmonds,B.T., Micanovic,R., Ou,W., Su,E.W., Tschang,S.H. and
            Wang,H.
  TITLE     Novel secreted proteins and their uses
  JOURNAL   Patent: WO 0214358-A 29 21-FEB-2002;
            ELI LILLY AND COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..445
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             25..333
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD29041.1"
                     /db_xref="GI:19701035"
                     /translation="MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAV
                     DCSWTLPPAPNSTSPVSFIATYRSESLEGGLRDTGLPGEPRTLALRALGSGKLETPPS
                     "
     mat_peptide     76..330
                     /product="unnamed"
BASE COUNT      124 a     138 c     115 g     68 t
ORIGIN
```

```
Query Match     23.6%;  Score 205.2;  DB 6;  Length 445;
Best Local Similarity   90.5%;  Pred. No. 1.8e-31;
Matches 219;  Conservative   0;  Mismatches 23;  Indels   0;  Gaps   0;
```

```
Qy    170 CATGACCCCGCAGCTTCTCCTGGCCCTTGTCCTCTGGGCCAGCTGCCCGCCCTGCGGTGG 229
          ||||||||||||||||| |||||||||||||||||||||||||||||| | |||| ||||
Db     24 CATGACCCCGCAGCTTCTCCTGGCCCTTGTCCTCTGGGCCAGCTGCCCGCCCTGCGGTGG 83

Qy    230 AAGGAAAAGGGCCCCCCAGCAGCTCTGACACTGCCCCGGGTGCAATGCCGAGCCGTTCGGTA 289
          |||||||||||| |||||||||||| |||||||| ||| |||||| | ||||| ||||||||
Db     84 AAGGAAAAGGGCCCCCCAGCAGCTCTGACACTGCCCCGGGTGCAATGCCGAGCCGTTCGGTA 143

Qy    290 CCCGGATCGCGGTGGATTGCTCCTGGACCCCTGCCTCCAAAACTCCACCAGCCCCGT 349
          |||||||||| ||||||||||||||||||| |||| ||||||||||||||||| |||
Db    144 CCCGGATCGCGGTGGATTGCTCCTGGACCCTGCCTCCAAACTCCACCAGCCCCGT 203

Qy    350 GTCCTTCATTGCCACGTACAGGGCTCGGCATGGCTGCCGGGGGCCACACCTGGCCTTGCCT 409
          ||||||||||||||| |||| || |  || |||||| ||||||| ||||| ||| |||
Db    204 GTCCTTCATTGCCACGTACAGGTCGGAGAGCCTGGAAGGGGGACACTGGACT 263

Qy    410 GC 411
          ||
Db    264 TC 265
```

```
RESULT 9
AC005578
```

```
LOCUS       AC005578               43297 bp    DNA     linear   PRI 01-SEP-1998
DEFINITION  Homo sapiens chromosome 19, cosmid F20887, complete sequence.
ACCESSION   AC005578
VERSION     AC005578.1  GI:3510237
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43297)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,J., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankhein,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 2.5 Mb region in 19p13.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43297)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Cosmid
            F20887 overlaps cosmid R33243 to the left by at least 8 kb and
            overlaps BC84616 (CIT-B-144d21) to the right from bases 41,733 to
            43,297. Additional chromosome 19 map and sequence information may
            be obtained at: http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1..43297
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19p13.3 between D19S883 and D19S325"
                     /clone="F20887"
                     /cell_line="UV5HL9-5B"
                     /clone_lib="LL19NC02 F chromosome 19-specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from human-hamster hybrid UV5HL9-5B, which
                     carries chromosome 19 as its only human chromosome."
     CDS             join(<177..283,457..644,1440..1477,2941..3048,5246..5374,
                     8654..8705,8927..9078,9226..9936,12267..12434,
                     16803..16868,17104..17181)
                     /note="Hypothetical human protein most similar to PID
                     g1429314 and several ankyrin-like proteins"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F20887_1, partial CDS"
                     /protein_id="AAC33490.1"
                     /db_xref="GI:3510240"
                     /translation="SGATPLIIAAQMCHTDLCRLLLQQGAAANDQDLQGRTALMLACE
                     GASPETVEVLLQGGAQPGITDALGQDAAHYGALAGDKLILHLLQEAAQRPSPSALTE
                     DDSGEASSQDDRDAYEBIVRLROEGRLLQKIRGLEQHKERRQQEROLSLHCCSQVQE
                     LQQLIVEROEKKESLGREVESLQSRLSLLENERENTISYDVTTLQDEEGPQGLTLRPPP
                     GAEVLLSRQLSPSAQEHLASLQEQVAVLTRQNOELMEKVQLLENFEKDETQMEVERLA
                     EVIPLALYDSLRAEFDLRQHAEALLQALRQQETREVPREEGAACGESEVAGATATKN
                     GPTHMELNGSVAPETKVNGAETIDEEAAGDETMEARTMEAEATGAEATGAEATGAKVT
                     ETKPTGAEVREMETTEEEANMETKPTGAQADTETTVGAVEATKTKAEEAEMQAY
                     GVGAQAEPPVTGTTNMEATGSRATGMESTGVSATGVENPGVEATVQREALFMKSERH
                     AAEAQLATAEQQLRGLRTEAERARQASRAQEALDKAKEKDKKEAARDHSSVVALYRS
                     HLLYA1QGQMDEDVQRILSQILQMQRLQAQGR"
     repeat_region   318..346
                     /rpt_family="MIR"
     misc_feature    457..644
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 93.000"
     repeat_region   complement(1043..1258)
                     /rpt_family="AluSq"
     misc_feature    1440..1477
                     /note="predicted exon, program: grail2exons_human_1.3,
```

                     frame: 1, quality: good, score: 68.000"
     repeat_region   complement(1785..2083)
                     /rpt_family="AluSx"
     repeat_region   complement(2124..2421)
                     /rpt_family="AluSx"
     misc_feature   2734..2814
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 80.000"
     misc_feature   2941..3048
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 96.000"
     repeat_region   3114..3264
                     /rpt_family="LINE2"
     repeat_region   complement(3505..3808)
                     /rpt_family="AluSq"
     repeat_region   4033..4330
                     /rpt_family="AluSx"
     repeat_region   4495..4537
                     /rpt_family="LINE2"
     repeat_region   4550..4587
                     /rpt_family="LINE2"
     repeat_region   4595..4894
                     /rpt_family="AluY"
     repeat_region   4968..5001
                     /rpt_family="LINE2"
     misc_feature   5246..5374
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 99.000"
     repeat_region   5527..5817
                     /rpt_family="AluSg"
     repeat_region   complement(5896..6085)
                     /rpt_family="(GGAA)n"
     repeat_region   complement(6086..6129)
                     /rpt_family="(GGAA)n"
     repeat_region   complement(6130..6323)
                     /rpt_family="AluSg/x"
     repeat_region   complement(6326..6608)
                     /rpt_family="AluSq"
     repeat_region   6700..6833
                     /rpt_family="(CA)n"
     repeat_region   complement(6845..7099)
                     /rpt_family="AluUb"
     repeat_region   7249..7545
                     /rpt_family="AluSx"
     repeat_region   7791..7879
                     /rpt_family="L1"
     repeat_region   7880..8157
                     /rpt_family="AluSx"
     repeat_region   8161..8464
                     /rpt_family="AluJo"
     misc_feature   8654..8705
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 91.000"
     misc_feature   8960..9087
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 100.000"
     misc_feature   9226..10839
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 92.000"
     repeat_region   complement(10973..11275)
                     /rpt_family="AluSx"
     repeat_region   complement(11334..11425)
                     /rpt_family="(GGAA)n"
     repeat_region   complement(11426..11678)
                     /rpt_family="AluJo"
     repeat_region   11680..11981
                     /rpt_family="AluSx"
     repeat_region   11982..12179
                     /rpt_family="AluUb"
     misc_feature   12267..12434
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 100.000"
     repeat_region   complement(12536..12569)

```
repeat_region       /rpt_family="MIR"
repeat_region       complement(12653..12945)
                    /rpt_family="AluSx"
repeat_region       complement(12957..13135)
                    /rpt_family="MIR"
repeat_region       complement(13229..13530)
                    /rpt_family="AluY"
repeat_region       complement(13580..13751)
                    /rpt_family="AluJ"
repeat_region       complement(13755..14043)
                    /rpt_family="AluSx"
repeat_region       complement(14074..14203)
                    /rpt_family="AluJ"
repeat_region       complement(14214..14318)
                    /rpt_family="MIR"
repeat_region       14358..14430
                    /rpt_family="MIR"
repeat_region       complement(14465..14667)
                    /rpt_family="MER20"
repeat_region       complement(14697..14731)
                    /rpt_family="LINE2"
repeat_region       14783..15083
                    /rpt_family="AluSx"
repeat_region       15087..15137
                    /rpt_family="MER58B"
repeat_region       15178..15228
                    /rpt_family="MER58B"
repeat_region       complement(15624..15908)
                    /rpt_family="AluSg"
repeat_region       complement(15519..16042)
                    /rpt_family="AluJo/FRAM"
repeat_region       16045..16077
                    /rpt_family="(TA)n"
repeat_region       complement(16078..16246)
                    /rpt_family="AluSx"
repeat_region       complement(16251..16539)
                    /rpt_family="AluSc"
repeat_region       complement(16540..16675)
                    /rpt_family="AluSx"
misc_feature        16603..16868
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 2, quality: excellent, score: 89.000"
repeat_region       16696..16955
                    /rpt_family="MIR"
repeat_region       16896..17181
                    /rpt_family="MIR"
misc_feature        17104..17181
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: excellent, score: 90.000"
repeat_region       complement(17728..17925)
                    /rpt_family="AluY"
repeat_region       complement(17926..18223)
                    /rpt_family="AluSx"
repeat_region       18327..18613
                    /rpt_family="AluJo"
repeat_region       complement(18910..19211)
                    /rpt_family="AluSx"
```

```
Query Match
Best Local Similarity   89.1%;          Score 182.6; DB 9;  Length 43297;
Matches 197; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  Pred. No. 3.7e-27;

Qy      334  ACTCCACCAGCCCCCGTCGTCCTTCATTGCCACGTACAGCGTCGGCGTGCCTGCCGGCCGGCGGCC  393
             |||||  | ||  | |  |   | |  |  | |||  |  ||  |  |||||  | ||||||||||||
Db    25768  ACTCCCTGAGGCGCTCAGGGAGCGCCCCACCCTGTTCGAGGCTCGGCGGCATGGCTGCCCGGGGGC  25827

Qy      394  ACAGCTGGCCTTGCTGCAGCAGACGACGCAAACGTCCACCAAGCTTCGACCATCACCGGATGTCC   453
             |||||||| || |||  | |||  ||  | || |||||||| |  |  | |||||||||||||||||
Db    25828  ACAGCTGGGCCCTGCCTGCACAGCAGCCAACGTCCACCAGCTGCACTGCCACCATCACCGGATGTCC  25887

Qy      454  AGCTGCGTTCTTCCCATGGCTCCCTACGTGCTCCAAATGTCACCGCCGTCCGCCCCCTGGGGCTCCA  513
             |||||||||||| ||||||||||||||||| || ||||||||||||||||||  ||||||||||||||
Db    25888  AGCTGCGTTCTCCCCATGGCTCCCTACGTCCTCAAATGTCACCGCCGTCCACCCCCTGGGGCTCCA   25947

Qy      514  GCAGCAGCTTCGTGCCTTTCATAACAGAGCACATCACTGAG  554
             |||||||||||||||||||||||||||||||||||||||||
Db    25948  GCAGCAGCTTCGTGCCTTTCATAACAGAGCACATCACTGAG  25988
```

```
Db    25948  GCAGCAGCTTCGTGCCTTTCATAACAGAGCACATCACTGAG  25988

RESULT 10
AC027272/c

LOCUS       AC027272               179607 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 19 clone RP11-486P13, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC027272
VERSION     AC027272.4  GI:9838224
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179607)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179607)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:8570236.
            Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----------  Project Information  ----------
            Center project name: H_NH0486P13
            ----------  Summary Statistics  ----------
            Sequencing vector: M13; 95%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 95% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 170867 bases at least Q40
            Consensus quality: 174124 bases at least Q30
            Consensus quality: 176036 bases at least Q20
            Insert size: 180000; agarose-fp
            Insert size: 179350; sum-of-contigs
            Quality coverage: 4.55 in Q20 bases; agarose-fp
            Quality coverage: 4.80 in Q20 bases; sum-of-contigs
            NOTE: This is a 'working draft' sequence. It currently
            consists of 18 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *        1     1338: contig of 1338 bp in length
            *     1339     1438: gap of unknown length
            *     1439     3330: contig of 1892 bp in length
            *     3331     3430: gap of unknown length
            *     3431     5429: contig of 1999 bp in length
            *     5430     5529: gap of unknown length
            *     5530     8107: contig of 2578 bp in length
            *     8108     8207: gap of unknown length
            *     8208    11180: contig of 2973 bp in length
            *    11181    11280: gap of unknown length
            *    11281    14039: contig of 2759 bp in length
            *    14040    14139: gap of unknown length
            *    14140    17230: contig of 3091 bp in length
            *    17231    17330: gap of unknown length
            *    17331    21334: contig of 4004 bp in length
            *    21335    21434: gap of unknown length
            *    21435    26877: contig of 5443 bp in length
            *    26878    26977: gap of unknown length
```

```
FEATURES             Location/Qualifiers
     source          1..179607
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-486P13"
     misc_feature    1..1338
                     /note="assembly_name:Contig8"
     misc_feature    1439..3330
                     /note="assembly_name:Contig9"
     misc_feature    3431..5429
                     /note="assembly_name:Contig10"
     misc_feature    5530..8107
                     /note="assembly_name:Contig11"
     misc_feature    8208..11180
                     /note="assembly_name:Contig12"
     misc_feature    11281..14039
                     /note="assembly_name:Contig13"
     misc_feature    14140..17230
                     /note="assembly_name:Contig14"
     misc_feature    17331..21334
                     /note="assembly_name:Contig15
                     vector_side:left"
     misc_feature    21435..26877
                     /clone_end:SP6
     misc_feature    26978..31453
                     /note="assembly_name:Contig16"
     misc_feature    3154..36515
                     /note="assembly_name:Contig17"
     misc_feature    36616..46678
                     /note="assembly_name:Contig18"
     misc_feature    46779..57999
                     /note="assembly_name:Contig19"
     misc_feature    58100..74742
                     /note="assembly_name:Contig20"
     misc_feature    74843..92504
                     /note="assembly_name:Contig21"
     misc_feature    92605..119545
                     /note="assembly_name:Contig22"
     misc_feature    119646..147254
                     /note="assembly_name:Contig23"
     misc_feature    147355..179607
                     /note="assembly_name:Contig24"

*          26978  31453: contig of 4476 bp in length
*          31454  31553: gap of unknown length
*          31554  36515: contig of 4962 bp in length
*          36516  36615: gap of unknown length
*          36616  46678: contig of 10063 bp in length
*          46679  46678: gap of unknown length
*          46679  57999: contig of 11221 bp in length
*          58000  58099: gap of unknown length
*          58100  74742: contig of 16643 bp in length
*          74743  74842: gap of unknown length
*          74843  92504: contig of 17662 bp in length
*          92505  92604: gap of unknown length
*          92605  119545: contig of 26941 bp in length
*          119546 119645: gap of unknown length
*          119646 147254: contig of 27609 bp in length
*          147255 147354: gap of unknown length
*          147355 179607: contig of 32253 bp in length.

BASE COUNT    41107 a   48085 c   48884 g   39812 t   1719 others
ORIGIN

Query Match
Best Local Similarity   19.8%;   Score 172.2;   DB 2;   Length 179607;
Matches  197;  Conservative   0;  Mismatches  23;  Indels  1;  Gaps  1;

Qy      334 ACTCCACCAGCCCCCCTCGTTCTTCATTGCCACGTACAGGCTCTCGGCATGGCTCCCGGGCC 393
            |||||  ||  |  ||||||  ||||||||||||||||||||||||||||||||||||||||
Db   158859 ACTCCCTGAGGCGGCTCAGCGAGCCCCACCTCTGTGCAGGGCTCGGCATGGCTCCCGGGGCC 158500

Qy      394 ACAGCTGGGCCTTGCCTGCCAGCAGACGCCAAGCGTCCACCACCACCATCACGGATGTCC 453
```

```
Db   158499 ACAGCTGGGCC TGCCTGCCAGCAGACGCCAACGGTCACCACGTCACCACCTGCACCATCACGGATGTCC 158441
            |||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      454 AGCTGTTCTCCATGGCTCCCTACGTGCCTCAATGTCACCGCCGCGCCCCCCTGGGGGGCTCCA 513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   158440 AGCTGTTCTCCATGGCTCCCTACGTGCCTCAATGTCACCGCCGCGCCCCCCTGGGGGGCTCCA 158381

Qy      514 GCAGCAGCTTCGTGCCTCTTTCATAACAGAGACACATCAAG 554
            ||||||||||||||||||||||||||||||||||||||||
Db   158380 GCAGCAGCTTCGTGCCTCTTTTCATAACAGAGACACATCAGTGAG 158340

RESULT 11
G54090
LOCUS       G54090         116 bp    DNA    linear    STS 15-JUL-1999
DEFINITION  902 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G54090
VERSION     G54090.1  GI:5499577
KEYWORDS    STS.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116)
  AUTHORS   Wong,G.K.S., Yu,J., Yang,Z., Hughes,W., Passey,D., Kibukawa,M.,
            Olson,V., Wilson,C. and Green,P.
  TITLE     Detection and Validation of Single Nucleotide Polymorphisms from
            EST Databases
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Gane Ka-Shu Wong
            Genome Center
            University of Washington
            Fluke Hall, Box 352145, Seattle, WA 98195, USA
            Tel: 206/685-7348
            Fax: 206/685-7344
            Email: gksw@u.washington.edu
            Primer A: CCGAGCCAGGTACTACACATCC
            Primer B: GAAGCCCTTGCTACTTGCC
            STS size: 116
            PCR Profile:
            Presoak: 92 degrees C for 1.00 minute
            Denaturation: 92 degrees C for 0.16 minute
            Annealing: 65 degrees C for 0.50 minute
            Polymerization: 72 degrees C for 1.00 minute
            PCR Cycles: 35
            Thermal Cycler: Perkin Elmer TC
            Protocol:
            Template:           7 ng
            Primer:             each 0.5 uM
            dNTPs:              each 100 uM
            Taq Polymerase:     0.025 units/ul
            Total Vol:          10 ul
            Buffer:
            MgCl2:       1.5 mM
            KCl:          50 mM
            Tris-HCl:     10 mM
            pH:           8.3.
FEATURES             Location/Qualifiers
     source          1..116
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14 cM"
                     /clone_lib="Human"
     STS             1..116
     primer_bind     1..20
     primer_bind     complement(98..116)
BASE COUNT    27 a   36 c   35 g   18 t
ORIGIN

Query Match
Best Local Similarity   11.8%;   Score 102.8;   DB 11;   Length 116;
                        98.1%;   Pred. No. 8.9e-11;
```

Matches 104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    755 CCGAGCCAGGTACTACGTCCAAGTGGGCGGCTCAGGACCTCACAGAACTACGGGGAACTGAG 814
          |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db      1 CCGAGCCAGGTACTACACATCCAAGTGGGCGGCTCAGGACCTCACAGACTACGGGGAACTGAG 60

Qy    815 TGACTGGGAGTCTCCCGCGCCACTGCCACAATGAGCCTGGGCAAGTAG 860
          ||||||||||||||||||||||||||||||| ||||||||||||||||
Db     61 TGACTGGGAGTCTCCCGCGCCACTGCCACAGATGAGCCTGGGCAAGTAG 106
```

RESULT 12
LOCUS   AC073737   193355 bp   DNA   linear   HTG 18-JUL-2000
DEFINITION Mus musculus clone RP23-264G1, WORKING DRAFT SEQUENCE, 24 ordered
            pieces.
ACCESSION   AC073737
VERSION   AC073737.2 GI:9256775
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE   Mus musculus
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 193355)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 193355)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Jul 18, 2000 this sequence version replaced gi:8810354.
            * -----------
            * Genome Center
            * Center: Joint Genome Institute
            * Center Code: JGI
            * Web site: http://www.jgi.doe.gov
            * -----------
            * Project Information
            * Center Project Name: 1838393
            * Center clone name: RPCI-23_264G1
            * -----------
            * Summary Statistics
            * Consensus quality: 175100 bases at least Q40
            * Consensus quality: 187679 bases at least Q30
            * Consensus quality: 188929 bases at least Q20
            * Estimated insert size: 87000; agarose-fp estimation
            * Estimated insert size: 192255; sum-of-contigs estimation
            * Quality coverage: 10.37 in Q20 bases; agarose-fp estimation
            * Quality coverage: 4.69 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *
            *   1   10615: contig of 10615 bp in length
            *   10616   10715: gap of unknown length
            *   10716   40728: contig of 30013 bp in length
            *   40729   40828: gap of unknown length
            *   40829   43385: contig of 2557 bp in length
            *   43386   43485: gap of unknown length
            *   43486   60463: contig of 16978 bp in length
            *   60464   60563: gap of unknown length
            *   60564   63757: contig of 3194 bp in length
            *   63758   63857: gap of unknown length
            *   63858   72484: contig of 8627 bp in length
            *   72485   72584: gap of unknown length
            *   72585   80251: contig of 7667 bp in length
            *   80252   80351: gap of unknown length
            *   80352   101558: contig of 21207 bp in length
            *   101559   101658: gap of unknown length
            *   101659   113941: contig of 12283 bp in length
            *   113942   114041: gap of unknown length
            *   114042   118302: contig of 4261 bp in length
            *   118303   118402: gap of unknown length
            *   118403   122141: contig of 3739 bp in length
            *   122142   122241: gap of unknown length
            *   122242   141898: contig of 19657 bp in length
            *   141899   141998: gap of unknown length
            *   141999   145605: contig of 3607 bp in length
            *   145606   145705: gap of unknown length
            *   145706   148555: contig of 2850 bp in length
            *   148556   148655: gap of unknown length
            *   148656   151425: contig of 2770 bp in length
            *   151426   151525: gap of unknown length
            *   151526   158286: contig of 6761 bp in length
            *   158287   158386: gap of unknown length
            *   158387   160638: contig of 2252 bp in length
            *   160639   160738: gap of unknown length
            *   160739   173388: contig of 12650 bp in length
            *   173389   173488: gap of unknown length
            *   173489   179077: contig of 5589 bp in length
            *   179078   179177: gap of unknown length
            *   179178   184518: contig of 5341 bp in length
            *   184519   184618: gap of unknown length
            *   184619   189175: contig of 4557 bp in length
            *   189176   189275: gap of unknown length
            *   189276   191154: contig of 1879 bp in length
            *   191155   191254: gap of unknown length
            *   191255   192103: contig of 849 bp in length
            *   192104   192203: gap of unknown length
            *   192204   193355: contig of 1152 bp in length.
FEATURES   Location/Qualifiers
     source   1..193355
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-264G1"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT   49405 a 44453 c 45086 g 52104 t 2307 others
ORIGIN
```

RESULT 13
LOCUS       AC118074/c            176359 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-47J17, *** SEQUENCING IN PROGRESS
            ***, 69 unordered pieces.
ACCESSION   AC118074
VERSION     AC118074.4  GI:21746302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus

Query Match 11.2%;  Score 97;  DB 2;  Length 193355;
Best Local Similarity 68.9%;
Matches  133; Conservative  0; Mismatches  60; Indels  0; Gaps  0;

```
Qy    362 CACGTACAGGCGTCGGCATGGCTGTCGCCCGGGGGCCCAGCGTGGCCTGCCTGCCTGCAGCAGACGCC 421
          |  | ||  | | ||| ||  |  ||  |||| | ||||| ||||| ||||||||  |||  || |
Db 111878 CCCATACGCAGCGCTCGGTGTGGCCACCAGACAGGAGGCCAGCCCCTGCCTACCTTGCCTACCTGCAC 111819

Qy    422 AACGTCCACCAGCTCCACATCACGGAGGATGTCCAGCTGTTCTCCATGGGTCCCTACGTGCT 481
          ||  || ||| |  ||  | || ||  |  | | ||  | ||  | |  ||  ||  ||| ||
Db 111818 CCAGGCCTCCGATGGCACCATCCCCGACAGTGCACCTGTTCCCACGGGTGCCTACCACATGCT 111759

Qy    482 CAATGTCACCGCGCCGTCCGCCCTGGGCTCCAGCAGCAGCTTCGTTGCCGTTTCATAAACAGA 541
          |||  | ||| | || |  ||||| ||| || ||| || || | | || ||  || |  | |
Db 111758 AAATGTCACTGCAGTGCCACCCCAGCGGCGGCGCCAGCAGCAGCGCCTCCTAGCCTTTGTGGCTGA 111699

Qy    542 GCACATCATCAAG 554
          |  | ||| |
Db 111698 GCGAATCAGTGAG 111686
```

```
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE   1  (bases 1 to 176359)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C.C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,R.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,M., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,J., Nguyen,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Puiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Soderberg,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,L., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 176359)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (13-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 176359)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 14, 2002 this sequence version replaced gi:20177536.
            --------------    Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------    Project Information
            Center project name: GUEB
            Center clone name: CH230-47J17
            --------------    Summary Statistics
            Sequencing vector: plasmid;
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 11372 bases at least Q40
            Consensus quality: 121114 bases at least Q30
```

```
          --------------
          Consensus quality: 124422 bases at least Q20
          --------------
* NOTE: Estimated insert size may differ from sequence length
*       (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
  be preserved.
*
*        1  1219: contig of 1219 bp in length
*     1220  1319: gap of unknown length
*     1320  2744: contig of 1425 bp in length
*     2745  2844: gap of unknown length
*     2845  4066: contig of 1222 bp in length
*     4067  4166: gap of unknown length
*     4167  5794: contig of 1628 bp in length
*     5795  5894: gap of unknown length
*     5895  6936: contig of 1042 bp in length
*     6937  7036: gap of unknown length
*     7037  8767: contig of 1731 bp in length
*     8768  8867: gap of unknown length
*     8868 10546: contig of 1679 bp in length
*    10547 10646: gap of unknown length
*    10647 12271: contig of 1625 bp in length
*    12272 12371: gap of unknown length
*    12372 13483: contig of 1112 bp in length
*    13484 13583: gap of unknown length
*    13584 14951: contig of 1368 bp in length
*    14952 15051: gap of unknown length
*    15052 16270: contig of 1219 bp in length
*    16271 16370: gap of unknown length
*    16371 17804: contig of 1434 bp in length
*    17805 17904: gap of unknown length
*    17905 19394: contig of 1490 bp in length
*    19395 19494: gap of unknown length
*    19495 20712: contig of 1218 bp in length
*    20713 20812: gap of unknown length
*    20813 22799: contig of 1987 bp in length
*    22800 22899: gap of unknown length
*    22900 24177: contig of 1278 bp in length
*    24178 24277: gap of unknown length
*    24278 25821: contig of 1544 bp in length
*    25822 25921: gap of unknown length
*    25922 27784: contig of 1863 bp in length
*    27785 27884: gap of unknown length
*    27885 28934: contig of 1050 bp in length
*    28935 29034: gap of unknown length
*    29035 30354: contig of 1320 bp in length
*    30355 30454: gap of unknown length
*    30455 31701: contig of 1247 bp in length
*    31702 31801: gap of unknown length
*    31802 33337: contig of 1536 bp in length
*    33338 33437: gap of unknown length
*    33438 34728: contig of 1291 bp in length
*    34729 34828: gap of unknown length
*    34829 36591: contig of 1763 bp in length
*    36592 36691: gap of unknown length
*    36692 38003: contig of 1312 bp in length
*    38004 38103: gap of unknown length
*    38104 40127: contig of 2024 bp in length
*    40128 40227: gap of unknown length
*    40228 41691: contig of 1464 bp in length
*    41692 41791: gap of unknown length
*    41792 43535: contig of 1744 bp in length
*    43536 43635: gap of unknown length
*    43636 45466: contig of 1831 bp in length
*    45467 45566: gap of unknown length
*    45567 47021: contig of 1455 bp in length
*    47022 47121: gap of unknown length
*    47122 48801: contig of 1680 bp in length
```

```
*  48802  48901: gap of unknown length
*  48902  49935: contig of 1034 bp in length
*  49936  50035: gap of unknown length
*  50036  51484: contig of 1449 bp in length
*  51485  51584: gap of unknown length
*  51585  53366: contig of 1782 bp in length
*  53367  53466: gap of unknown length
*  53467  55706: contig of 2240 bp in length
*  55707  55806: gap of unknown length
*  55807  57484: contig of 1678 bp in length
*  57485  57584: gap of unknown length
*  57585  59524: contig of 1940 bp in length
*  59525  59624: gap of unknown length
*  59625  61504: contig of 1880 bp in length
*  61505  61604: gap of unknown length
*  61605  63038: contig of 1434 bp in length
*  63039  63138: gap of unknown length
*  63139  65148: contig of 2010 bp in length
*  65149  65248: gap of unknown length
*  65249  68165: contig of 2917 bp in length
*  68166  68265: gap of unknown length
*  68266  70437: contig of 2172 bp in length
*  70438  70537: gap of unknown length
*  70538  72778: contig of 2241 bp in length
*  72779  72878: gap of unknown length
*  72879  74865: contig of 1987 bp in length
*  74866  74965: gap of unknown length
*  74966  78152: contig of 3187 bp in length
*  78153  78252: gap of unknown length
*  78253  80830: contig of 2578 bp in length
*  80831  80930: gap of unknown length
*  80931  83775: contig of 2845 bp in length
*  83776  83875: gap of unknown length
*  83876  86764: contig of 2889 bp in length
*  86765  86864: gap of unknown length
*  86865  90526: contig of 3662 bp in length
*  90527  90626: gap of unknown length
*  90627  92648: contig of 2022 bp in length
*  92649  92748: gap of unknown length
*  92749  96414: contig of 3666 bp in length
*  96415  96514: gap of unknown length
*  96515  99630: contig of 3116 bp in length
*  99631  99730: gap of unknown length
*  99731 102795: contig of 3065 bp in length
* 102796 102895: gap of unknown length
* 102896 105478: contig of 2583 bp in length
```

RESULT 14

Query Match   10.4%; Score 90.2; DB 2; Length 176359;
Best Local Similarity 67.0%; Pred. No. 1.3e-08; DB 2;
Matches 128; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy    364 CGTACAGGCTCGGGCATGGCTGCGGGCCCGGGGGGCTGGCCTTGGCCCGGGCAGCACAGTGG 423
          |||| ||||| ||  ||| | |||| |||| |||||| |||||||  ||| ||||| |||
Db  66695 CATGCAGGCTGCGGTGTGGGCCGCGCCCAGCAGGAGAGCCAGCCTGCCTGCCAACGGACGACCCCC 66636

Qy    424 CGTCACCCAGCTGCACCAACACGGATGTCCAGCGTGTTCTCCATCGGCTCCCTACGTGCTCTCA 483
          ||| ||| ||||||| |||| |  ||  |||| |||||| ||| |||| |||||||  |||| ||
Db  66635 AGGCACCGCTTGCACCATCCCACAGCGTGCACCGTGTTTTCCACAGTGCCCTACTTGCTAA 66576

Qy    484 ATGTCACCGCCGCGTCCGCCCCTGGGGGGTCCAGCAGGCAGCAGCTTCGTGCGCCTTTCATAACAGACC 543
          ||||||| |  |  | |||||  |||| |||| |||||| ||||| | ||| |  ||| ||
Db  66575 ATGTCACTGTGAGTGCACCCTGGGGCGGGCCAACAGCAGCGCCTTCTAGCCTTTGTGTGCGGAAC 66516

Qy    544 ACATCACTCACAG 554
          |||| |||
Db  66515 GAATCAGTAAG 66505
```

ACCESSION   AX399177
DEFINITION Sequence 17 from Patent WO0194605.
LOCUS   AX399177   1332 bp   DNA   linear   PAT 27-MAY-2002

RESULT 15

VERSION   AX399177.1 GI:21261511
KEYWORDS
SOURCE   Rattus sp.
  ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Zolotukhin,S., Kalra,S.P., Moldawer,L.L., Muzyczka,N.,
            Hauswirth,W.W. and Kalra,P.S.
  TITLE   Aav vector-based compositions and therapeutic uses thereof
  JOURNAL   Patent: WO 0194605-A 17 13-DEC-2001;
            University of Florida (US)
FEATURES   Location/Qualifiers
     source   1..1332
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
BASE COUNT   291 a   428 c   338 g   275 t
ORIGIN

Query Match   8.6%; Score 74.8; DB 6; Length 1332;
Best Local Similarity 52.1%; Pred. No. 2.9e-05; DB 6;
Matches 219; Conservative 0; Mismatches 192; Indels 9; Gaps 2;

```
Qy    429 ACCAGGTGCACCATCACGGATGTCCAGCGTGTTCTCCATCGGCTCCCTACGTGCTCAATGTC 488
          |||  |||  |||| |||  |   ||||||||||| | |    |  ||   |||  |    ||
Db    564 AACCGCTGTCACATTCGGTACATGCACCGTGTTCTCAAACCATCAACGTACACAGGTCTCCATA 623

Qy    489 ACCGCCGTCCGCGCCCCTGGGGGGTCCAGCAGCAGCAGCTTCGTGCCTTTCATAACAGACAATC 548
          |   |||  | ||   | |||| | |||| ||  |    |  |    ||  | |||  |  |||
Db    624 AGTGTCACGAACGCCTTGGGTCACACAACCACCGGCTATCACCTTCGACGAATTCACCATT 683

Qy    549 ATCAAGCACCCGACCCTCCAGAAGGCGTG------CGCCTAAGCCCCCTCCGCGTGAGCGCCAG 602
          |||||||  ||||| |||||||||  ||       ||  || |     |||| || |  ||  |
Db    684 GTGAAGCCCGATCCTCCAGAAAATGTGGGTGGCCCGGGCAGTGCCGCCAGCACAACCCCGTCGA 743

Qy    603 CTACAGGTGCAGTGGGGAGCCTCCCGGGCCCTTCCCGCCCTTCCCAGAGATCTTCTCACTGAAG 662
          ||  |||||  | | |  | |||   |||  || ||  || ||  | || ||  | ||||||||
Db    744 CTGGAGGTGACATGGCAGAGACCCCCTCAAACTTGGCCTTCGATCCGGAATCCTTTCCACTCAAG 803

Qy    663 TACTGGATCCGTTACAACGTGCAGGGAGCTGCGCGCTTCCACCGGGGTGGGGCCCATTGAA 722
          |  |  ||||  ||||  |||     | ||  |  || || |  |  | ||  ||  | ||
Db    804 TTTTTTCTGCGTACCGGCCCTCATCCTGGATCAATGGCACACCATGTGACCGTCTCGAAT 863

Qy    723 GCCACGTCCTTCAATCCTCCAGGGCTGTGTCGCGGCCCCGAGCGCCCAGGTGCCCCAGTTGAA 782
          ||||| ||   | || || | ||  || |   || ||   |  |  ||   |  |  | | ||
Db    864 GGCACAGCCCACACCATCACGGATGCCTATTGCTGGGAAGGAGAGGAGTACTACACGGTGGCC 923

Qy    783 GCTCAGGACCTACAGAGACTACGGGGAACTGAGTGGACTGAGAGTCTCCCGCGCCACTGCCACA 842
          || ||| | ||||  | || ||  |  | |||| | ||||| |||  |  | ||||| |||||||
Db    924 GCCAAGGA---CAATGAGATTGCGACATGGAGTGACTGCGGAGTGTGGTCTGCTCACGCGCCACA 980
```

ACCESSION   S54212
VERSION   S54212.1 GI:265055
KEYWORDS
SOURCE   Rattus sp. brain.
  ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 1332)
  AUTHORS   Ip,N.Y., McClain,J., Barrezueta,N.X., Aldrich,T.H., Pan,L., Li,Y.,
            Wiegand,S.J., Friedman,B., Davis,S. and Yancopoulos,G.D.
  TITLE   The alpha component of the CNTF receptor is required for signaling
            and defines potential CNTF targets in the adult and during
            development
  JOURNAL   Neuron 10 (1), 89-102 (1993)

RESULT 15
S54212
LOCUS   S54212   1332 bp   mRNA   linear   ROD 08-MAY-1993
DEFINITION S54212 ciliary neurotrophic factor receptor alpha component [rats, brain, mRNA, 1332 nt].

```
MEDLINE     93152175
PUBMED      8381290
REMARK      GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 124437] from the original journal article.
            This sequence comes from Fig. 2A.
FEATURES             Location/Qualifiers
     source          1..1332
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
     gene            1..1332
                     /gene="ciliary neurotrophic factor receptor alpha
                     component, CNTFR alpha"
     CDS             81..1199
                     /gene="ciliary neurotrophic factor receptor alpha
                     component, CNTFR alpha"
                     /note="This sequence comes from Fig. 2A; CNTFR alpha"
                     /codon_start=1
                     /product="ciliary neurotrophic factor receptor alpha
                     component"
                     /protein_id="AAB25290.1"
                     /db_xref="GI:265056"
                     /translation="MAASVPWACCAVLAAAAAVYTQKHSPQEAPHVQYERLGTDVTL
                     PCGTASWDAAVTWRVNGTDLAPDLLNGSQLLIRSLELGHSGLYACPHRDSWHLRHQVL
                     LHVGLPPREPVLSCRSNTYPKGFYCSWHLSAPTYIPNTENVTVLHGSKMVCEKDPAL
                     KNRCHIRYMHLFSTIXYKVSISVSNALGHNTTAITFDEFTIVKPDPPENVARPVPSN
                     PRLEVTWQTPSTWPDPESFPLKFFLRYRPLILDQWQHVELSNGTAHTIIDAYAGKEY
                     IIQVAAKDNEIGTWSDWSVAAHATPWTEEPRHLTTEAQAPETTISTTSTSSLAPPTTKI
                     CDPGELSSGGPSIPFLTSVPVTLVLAAAAATANNLLI"
BASE COUNT       291 a      428 c      338 g      275 t
ORIGIN
```

```
Query Match            8.6%;   Score 74.8;   DB 10;   Length 1332;
Best Local Similarity 52.1%;   Pred. No. 2.9e-05;
Matches  219;  Conservative    0;  Mismatches 192;  Indels  9;  Gaps  2;

Qy   429 ACCAGCTGCACCATCACGGATGTCCAGCTGTTCTCCATGGCTCCCTACGTGCTCAATGTC 488
         | ||| |  | ||  |  ||  ||| ||| ||| || | || || | ||   | ||
Db   564 AACCGCTGTCACATTCGTACATGCACCTGTTCTCAACCATCAAGTACAAGGTCTCCATA 623

Qy   489 ACCGCCCGTCCGCCCCTGGGGCTCCAGCAGCTTCGTGCCTTTCATAACAGAGCACATC 548
         | ||| |  | |  ||  | ||   |||  || || | || |  ||
Db   624 AGTGTCAGCACACGCCTTGGGTCACAACACCACGGCTATCACCTTCGACGAATTCACCATT 683

Qy   549 ATCAAGCCCGACCCTCCAGAAGGCGTG---------CGCCTAAGCCCCTCCGCTGAGCGCCCAG 602
         ||| |   |||||||| |||| ||| |         ||| ||  |||  | |||  | ||
Db   684 GTGAAGCCCGATCCTCCAGAAAATGTGGCGCCCGGCCCAGTGCCCAGCAACCCCGTCGA 743

Qy   603 CTACAGCGTGCAGTGGGAGGCCTCCGGGGTCCTGGCCCCTTCCCAGAGATCTTCTCACTGAAG 662
         || ||  ||| || ||| || || ||| || ||  ||   |||||  || | || | ||
Db   744 CTGGAGGTGACATGGCAGAGCCCCTCAACTTGGCCTGATCCCGAATCCTTTCCACTCAAG 803

Qy   663 TACTGGAATCCGTTACAAGCGTCAGGGAGCTCCGCGCCGCTTCCACCGGGTGGGGCCATTGAA 722
         || | ||  ||||| |||||   | || |||| ||  |||  || | || |||  |||  ||
Db   804 TTTTTTCTGCGCTACCGGCCCTCTCATCCTGGATCAATGGCAGCATGTGGAGCTCTCGAAT 863

Qy   723 GCCACGTGCTTCATCCTCAGGGCTGTGCGGCCCCGAGCCACGGTACTACGTCCAAGTGGCG 782
         || ||| |||  | || |||| || | ||| | || |  ||   || ||| ||| || ||
Db   864 GGCACAGCCCACACCATCACGGATGCCCTATGCTGGGAAGAGTACATCATCCAAGTGGCC 923

Qy   783 GCTCAGGACCTCACAGACTACGGGGAAACTGAGTGACTGGGAGTCTCCCCGCCACTGCCACA 842
         ||  || ||| || |  ||  |  || | ||||| |  |||||||||  | || | ||||
Db   924 GCCAAGGA---CAATGAGATTGGGACATGGAGTGACTGGAGTGTGGCTGCTCACCCACA 980
```

```
Search completed: December 28, 2002, 09:42:41
Job time : 3429 secs
```

-
-
-
-

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: December 28, 2002, 10:16:24 ; Search time 1716.5 Seconds
(without alignments)
3255.309 Million cell updates/sec

Title: US-09-921-994-2
Perfect score: 1064
Sequence: 1 MTPQLLLALVLWASCPPRSG..........HNRAHHQARPSRRRAPKPPR 192

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV-xlp
-O=/cgn2_1/USPTO_spool/US09921994/runat_27122002_084132_6883/app_query.fasta_1.782
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09921994 @CGN_1_1_3637 @runat_27122002_084132_6883 Sequence 3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pri:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_Om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

..........HNRAHHQARPSRRRAPKPPR 192

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%	Query Match Length	DB	ID	Description
1	1064	100.0	868	6	AX377958	AX377958 Sequence
2	784.5	73.7	1195	9	BC015364	BC015364 Homo sapi
3	782	73.5	1161	6	AR003351	AR003351 Sequence
4	782	73.5	1161	6	AR050784	AR050784 Sequence
5	782	73.5	1161	6	L08187	L08187 Human cytok
6	374.5	35.2	43297	9	HUMEBI3X	HUMEBI3X
7	373	35.1	445	2	AC005578	AC005578 Homo sapi
8	351.5	33.0	1131	2	AX392987	AX392987 Sequence
9	281.5	26.5	179607	10	AC027272	AC027272 Homo sapl
10	281.5	26.5	1131	10	BC008209	BC008209 Mus muscu
11	193.5	18.2	193355	2	AF013114	AF013114 Mus muscu
12	173.5	16.3	176359	2	AC073737	AC073737 Mus muscu
13	149.5	14.1	10770	1	AC118074	AC118074 Rattus no
14	149	14.0	168964	2	AE012121	AE012121 Xanthomon
15	149	14.0	186558	2	AC131212	AC131212 Homo sapi
16	148	13.9	1969	2	AC079031	AC079031 Homo sapi
17	146.5	13.8	82746	7	I93634	I93634 Sequence 3
18	146	13.7	133330	2	AF453501	AF453501 Actinosyn
19	145.5	13.7	22449	2	OSJN00266	OSJN00266 Oryza sat
20	145.5	13.7	22449	2	SPSNBCDE	SPSNBCDE S.pristinae
21	145	13.6	1850	9	ACYFIMBA	ACYFIMBA A.viscosus
22	145	13.6	1969	6	M32067	M32067 A.viscosus
23	145	13.6	1969	6	I93632	I93632 Sequence 1
24	145	13.6	2664	8	2MAHAS108	2MAHAS108
25	144	13.5	38390	1	SC2H12	SC2H12 Streptomy
26	143.5	13.5	66618	9	HSJ579N16	HSJ579N16 Human DNA
27	143	13.4	132914	2	CNS08CAA	CNS08CAA
28	142.5	13.4	125505	2	AP004177	AP004177 Oryza sat
29	142	13.3	187757	2	AC087557	AC087557 Mus muscu
30	141.5	13.3	169230	9	AP002505	AP002505 Homo sapi
31	141.5	13.3	188604	9	AP002439	AP002439 Homo sapi
32	141	13.3	14472	1	AE003928	AE003928 Xylella f
33	141	13.3	18959	1	AE004032	AE004032 Xylella f
34	141	13.3	20956	1	AE004082	AE004082 Xylella f
35	140.5	13.2	135789	8	AC051634	AC051634 Oryza sat
36	140	13.2	4257	6	AR144708	AR144708 Sequence
37	140	13.2	4257	6	AR145616	AR145616 Sequence
38	140	13.2	6633	14	HEHSV1G3	HEHSV1G3
39	140	13.2	12001	6	AR048721	AR048721 Herpes simp
40	140	13.2	26245	14	L00036	L00036 Human herpe
41	140	13.2	26245	14	L00036	L00036 Human herpe
42	140	13.2	38425	1	SCF55	SCF55 Streptomy
43	140	13.2	152261	14	HEICG	HEICG Herpes simp
44	140	13.2	152261	14	HEICG	HEICG Herpes simp
45	139.5	13.1	1466	5	AF091043	AF091043 Coturnix

Result 1

RESULT 1

ALIGNMENTS

```
LOCUS       AX377958                 868 bp    DNA        linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0212282.
ACCESSION   AX377958
VERSION     AX377958.1  GI:19574012
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bowman,M.R.
  TITLE     Novel epi-3-alt protein and nucleic acid molecules and uses
            therefor
  JOURNAL   Patent: WO 0212282-A 1 14-FEB-2002;
            GENETICS INSTITUTE, INC. (US)
FEATURES             Location/Qualifiers
     source          1..868
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      147 a    325 c    239 g    157 t
ORIGIN
```

```
US-09-921-994-2 (1-192) x AX377958 (1-868)

Alignment Scores:
Pred. No.:          6.79e-34
Score:              1064.00
Percent Similarity: 100.00%        Length:       868
Percent Identity:   100.00%        Matches:      192
Best Local Similarity: 100.00%     Conservative: 0
Query Match:        6              Mismatches:   0
                                   Indels:       0
                                   Gaps:         0
```

```
Qy     1 MetThrProGlnLeuLeuLeuLeuAlaLeuValLeuTrpAlaSerCysProProArgSerGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGACCCCGCAGCTTCTCCTGCTGGCCCTTTGTCCTCCTGGGCCACGTTCCGGTAC 76

Qy    21 ArgLysGlyProProAlaAlaLeuThrLeuProArgValGlnCysArgAlaSerArgTyr 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    77 AGGAAAGGGCCCCCACCAGCTCTGACACTGCCCGCGGGTGCAATGCCGAGCCTCTCGGTAC 136

Qy    41 ProIleAlaValAspCysSerTrpThrLeuProHisAspCysArgAlaSerProGlyPro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 CCGATCGCCGTGGAATTGCTCCACGACTGCCGCAGCCCCAGCCTTTCTCCTGGCCCT 196

Qy    61 CysProLeuGlyGlnLeuProIleAlaLeuArgTrpLysGluArgAlaProSerSerSerAsp 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 TGTCCTCTGGGCCAGCTGCCCATCGCCGTGGAAGGAAAAGGGCCCCCAGCAGCTCTGAC 256

Qy    81 ThrAlaProGlyAlaMetProSerLeuSerValProAspArgArgGlyLeuLeuLeuAsp 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257 ACTGCCCCCGGGTGCAATGCCGAGCCTCTCGGTACCCGATCGCCGTGGAATTGCTCCTGGAC 316

Qy   101 ProAlaAlaCysSerLysLeuHisGlnProArgValLeuHisCysHisSerValGlnAlaArg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   317 CCTGCCCCTGCTCCAAACTCACACGCCGTCTCCTTCATTGCCACGTACAGGCTCGG 376

Qy   121 HisGlyCysProGlyProGlnLeuProIleAlaLeuAlaAlaAspAlaAsnValHisGlnLeu 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   377 CATGGCTGCCCGGGGCCACAGCTGCCCTTGCCTCGCAGACAGCAGCCGTCCACCAGCTG 436

Qy   141 HisHisHisGlyCysProAlaValLeuHisGlySerLeuArgAlaGlnCysProHisSerArgArg 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   437 CACCATCACGGATGTGCCAGCTGTTCTCCATGGCTCCCTACGGTGCTCAAATGTCACGGCCCGT 496

Qy   161 ProProLeuGlyLeuGlnLeuGlnLeuArgAlaAlaPheHisAsnArgAlaHisHisGlnAla 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   497 CCGCCCCTTGGGGGCCTCCAGCAGCTTCGTTGTCCCTTTCATTAACACAGAGCACATCATCAAGCC 556

Qy   181 ArgProSerArgArgAlaProLysProProArg 192
           |||||||||||||||||||||||||||||||
Db   557 CGACCCTCCAGAAGGCCGTGCGCCTAAGCCCCCTCGC 592
```

```
RESULT 2
LOCUS       BC015364                1195 bp    mRNA       linear   PRI 04-OCT-2001
DEFINITION  Homo sapiens, Epstein-Barr virus induced gene 3, clone MGC:21515
            IMAGE:3897136, mRNA, complete cds.
ACCESSION   BC015364
VERSION     BC015364.1  GI:15929884
KEYWORDS    MGC.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1195)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site:  http://www-shgc.stanford.edu
            Contact:   (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 22 Row: f Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1..1195
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:21515 IMAGE:3897136"
                     /tissue_type="Pancreas, epithelioid carcinoma"
                     /clone_lib="NIH_MGC_70"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             54..743
                     /codon_start=1
                     /product="Epstein-Barr virus induced gene 3"
                     /protein_id="AAH15364.1"
                     /db_xref="GI:15929885"
                     /translation="MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAV
                     DCSWTLPDAPNSTSPVSFIATYRLGMAARGHSWPGLQQTPTSTSCTITDVQLFSMAPY
                     VLNVTAVHPWGSSSSFVPFITEHIIKPDPPEGVRLSPLAEROLQVQWEPPGSWPFPEI
                     FSLKYWIRYKRQGAARHRVGPIEAFSFILRAVRPRARYYQVAAQDLTDYGELSDWS
                     LPATATMSLGK"
BASE COUNT      247 a    395 c    311 g    242 t
ORIGIN
```

```
US-09-921-994-2 (1-192) x BC015364 (1-1195)

Alignment Scores:
Pred. No.:          5.94e-23
Score:              784.50
Percent Similarity: 90.12%         Length:       1195
Percent Identity:   88.89%         Matches:      144
Best Local Similarity: 73.73%      Conservative: 2
Query Match:        9              Mismatches:   5
                                   Indels:       11
                                   Gaps:         1
```

```
Qy    31 ProArgValGlnCysArgAlaSerArgTyrProIleAlaValAspCysSerTrpThrLeu 50
           |||                                   :::            |||||
```

```
Db   23  CCCACTCCTGAGCAGCAGAGCTGGCGCGC----------                         49
Qy   51  ProHisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArg    70
Db   50  AGCCATGACCCGCAGCTTCTCCTGGCCCTTGTCCTTCTGGCCAGCTGCCCGCCCTGCAG    109
Qy   71  TrpLysGluArgAlaProSerSerSerAspThrAlaProGlyAlaMetProSerLeuSer    90
Db   71  TGGAAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCCCCGGGTGCAATGCCGAGCCTCTCG   169
Qy   91  ValProAspArgArgGlyLeuLeuLeuAspProAlaAlaCysSerLysLysLeuHisGlnPro  110
Db   170 GTACCCGACCGGCGGGATTGCTCCTGGACCCCGCGGCTTGCTCCAAAACTCCACCAGCC    229
Qy   111 ArgValLeuHisCysHisHisValGlnAlaAlaArgHisGlyGlyCysProGlnProGlnLeuAlaLeu  130
Db   230 CGTGTCCTTCATTGCCACGGTACAGGCTGCCCGGGGGTGCCCACAGCTGGCCCTG        289
Qy   131 ProAlaAlaAlaSpAlaAsnValHisGlnLeuHisHisHisSerGlyCysProAlaValLeuHis  150
Db   290 CCTGCAGCAGACGCCAAACGTCCACCAGCTGCCACCATCACGGGATGTCCAGCTGTTCTCCAT  349
Qy   151 GlySerLeuArgGlyAlaGlnCysTyrHisLeuArgGlyArgProProLeuGlyLeuGlnGlyLeuArg  170
Db   350 GGCTCCCTACGTGTCTCAATGTCACCCGCCGTCCACCCCGTGGGGCTCCACCAGCAGCTTCGT  409
Qy   171 AlaPheHisAsnArgAlaHisHisGlnAlaArgProSerArgArgArgAlaProLysPro   190
Db   410 GCCTTCATAACAGACACCATCATCAAGCCCGACCCTCCAGAAGGCCGTGCCCTAAGCCC    469
Qy   191 ProArg  192
Db   470 CCCTCGC  475
```

RESULT 3

```
LOCUS       AR003351                1161 bp    DNA        linear   PAT 04-DEC-1998
DEFINITION  Sequence 5 from patent US 5744301.
ACCESSION   AR003351
VERSION     AR003351.1  GI:3964610
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Birkenbach,M. and Kieff,E.
  TITLE     Methods of detection of epstein barr virus induced genes expressed
            in the placenta
  JOURNAL   Patent: US 5744301-A 5 28-APR-1998;
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="unknown"
BASE COUNT    239 a    378 c    304 g    240 t
ORIGIN
```

US-09-921-994-2 (1-192) x AR003351 (1-1161)

Alignment Scores:
```
Pred. No.:
Score:                  7.24e-23
Percent Similarity:     782.00                  Length:         1161
Best Local Similarity:  100.00%                 Matches:         140
Query Match:            99.29%                   Conservative:      1
DB:                     73.50%                   Mismatches:        0
                        6                        Indels:            0
                                                 Gaps:              0
```

```
Qy   52  HisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArgTrp   71
Db   13  CATGACCCGCAGCTTCTCCTGGCCCTTGTCCTTCTGGCCAGCTGCCCG::::           72
Qy   72  LysGluArgAlaProSerSerSerAspThrAlaProGlyAlaMetProSerLeuSerVal   91
Db   73  AAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCCCCGGGTGCAATGCCGAGCCTCTCGGTA  132
```

```
Qy   92  ProAspArgArgGlyLeuLeuLeuAspProAlaAlaCysSerSerLysLysLeuHisGlnProArg  111
Db   133 CCCGACCGGCGGGATTGCTCCTGGACCCCGCGGCTTGCTCCAAAACTCCACCAGCCGT    192
Qy   112 ValLeuHisCysHisHisValGlnAlaAlaArgHisGlyGlyCysProGlnProGlnLeuAlaLeuPro  131
Db   193 GTCCTTCATTGCCACGGTACAGGCTGCCCGGGGGTGCCCACAGCTGGCCCTGCCT        252
Qy   132 AlaAlaAlaSpAlaAsnValHisGlnLeuHisHisHisSerGlyCysProAlaValLeuHisGly  151
Db   253 GCAGCAGACGCCAAACGTCCACCAGCTGCCACCATCACGGGATGTCCAGCTGTTCTCCATGGC  312
Qy   152 SerLeuArgGlyAlaGlnCysTyrHisLeuArgGlyArgProProLeuGlyLeuGlnGlyLeuArgAla  171
Db   313 TCCCTACGTGTCTCAATGTCACCCGCCGTCCACCCCGTGGGGCTCCACCAGCAGCTTCGTGCC  372
Qy   172 PheHisAsnArgAlaHisHisGlnAlaArgProSerArgArgArgAlaProLysProPro   191
Db   373 TTTCATAACAGACACCATCATCAAGCCCGACCCTCCAGAAGGCCGTGCCCTAAGCCCCCCT  432
Qy   192 Arg  192
Db   433 CGC  435
```

RESULT 4

```
LOCUS       AR050784                1161 bp    DNA        linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5830451.
ACCESSION   AR050784
VERSION     AR050784.1  GI:5974148
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Devergne,O. and Kieff,E.D.
  TITLE     Haematopoietic cytokine Epstein Barr virus-induced protein
  JOURNAL   Patent: US 5830451-A 1 03-NOV-1998;
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="unknown"
BASE COUNT    239 a    378 c    304 g    240 t
ORIGIN
```

US-09-921-994-2 (1-192) x AR050784 (1-1161)

Alignment Scores:
```
Pred. No.:
Score:                  7.24e-23
Percent Similarity:     782.00                  Length:         1161
Best Local Similarity:  100.00%                 Matches:         140
Query Match:            99.29%                   Conservative:      1
DB:                     73.50%                   Mismatches:        0
                        6                        Indels:            0
                                                 Gaps:              0
```

```
Qy   52  HisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArgTrp   71
Db   13  CATGACCCGCAGCTTCTCCTGGCCCTTGTCCTTCTGGCCAGCTGCCCG::::           72
Qy   72  LysGluArgAlaProSerSerSerAspThrAlaProGlyAlaMetProSerLeuSerVal   91
Db   73  AAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCCCCGGGTGCAATGCCGAGCCTCTCGGTA  132
```

```
Qy   92  ProAspArgArgGlyLeuLeuLeuAspProAlaAlaCysSerSerLysLysLeuHisGlnProArg  111
Db   133 CCCGACCGGCGGGATTGCTCCTGGACCCCGCGGCTTGCTCCAAAACTCCACCAGCCGT    192
Qy   112 ValLeuHisCysHisHisValGlnAlaAlaArgHisGlyGlyCysProGlnProGlnLeuAlaLeuPro  131
Db   193 GTCCTTCATTGCCACGGTACAGGCTGCCCGGGGGTGCCCACAGCTGGCCCTGCCT        252
Qy   132 AlaAlaAlaSpAlaAsnValHisGlnLeuHisHisHisSerGlyCysProAlaValLeuHisGly  151
Db   253 GCAGCAGACGCCAAACGTCCACCAGCTGCCACCATCACGGGATGTCCAGCTGTTCTCCATGGC  312
```

```
Qy  152  SerLeuArgArgAlaGlnCysHisArgArgArgProProLeuGlyLeuGlnGlnLeuArgArgAla  171
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313  TCCCTACGTGCTCAATGTCACCGCCGTCCACCCCTGGGGCTCCAGCAGCAGCTTCGTGCC  372

Qy  172  PheHisAsnArgAlaIaHisHisGlnInAlaArgProSerArgArgArgGlaAlaProLysProPro  191
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  373  TTTCATAAACAGAGCACATCATCAAGCCCGACCCTCCAGAAGGCGTGCGCCTTAAGCCCCCT  432

Qy  192  Arg  192
         |||
Db  433  CGC  435
```

RESULT 5

```
LOCUS       HUMEBI3X      1161 bp    mRNA    linear    PRI 01-APR-1996
DEFINITION  Human cytokine receptor (EBI3) mRNA, complete cds.
ACCESSION   L08187
VERSION     L08187.1  GI:632973
KEYWORDS    cytokine receptor.
SOURCE      Homo sapiens cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1161)
AUTHORS     Devergne,O., Hummel,M., Koeppen,H., Le Beau,M.M., Nathanson,E.C.,
            Kieff,E. and Birkenbach,M.
TITLE       A novel interleukin-12 p40-related protein induced by latent
            Epstein-Barr virus infection in B lymphocytes
JOURNAL     J. Virol. 70 (2), 1143-1153 (1996)
MEDLINE     96135230
PUBMED      8551575
COMMENT     On Jan 24, 1995 this sequence version replaced gi:181921.
FEATURES    Location/Qualifiers
     source  1..1161
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /cell_line="BL41/B95-8"
             /cell_type="B lymphocyte, EBV-converted Burkitt lymphoma"
     gene    1..1161
             /gene="EBI3"
     CDS     14..703
             /gene="EBI3"
             /codon_start=1
             /product="cytokine receptor"
             /protein_id="AAA93193.1"
             /db_xref="GI:632974"
             /translation="MTPQLLLALVLWASCPPCSGRKGGPPAALTLPRVQCRASRYPIAV
             DCSWTLPPAPNSTSPVSFIATYRLGMAARGHSWPCLQQTPTSTSCTITDVQLFSMAPY
             VLNVTAVHPWGSSSSFVPFITEHIIKPDPPEGVRLSPLAERIHVQVQWEPPGSWPFPEI
             FSLKYWIRYKRQGAARFHRVGPIEATSFILRAVRPRARYYVQVAAQDLTDYGELSDWS
             LPAIATMSLGK"
     sig_peptide  14..73
                  /gene="EBI3"
     mat_peptide  74..700
                  /gene="EBI3"
                  /product="cytokine receptor"
BASE COUNT    239 a   378 c   304 g   240 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:
Score:                     7.24e-23
                           782.00          Length:          1161
Percent Similarity:        100.00%         Matches:         140
Best Local Similarity:     99.29%          Conservative:    1
Query Match:               73.50%          Mismatches:      1
DB:                        9               Indels:          0
                                           Gaps:            0
```

US-09-921-994-2 (1-192) x HUMEBI3X (1-1161)

```
Qy  52  HisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArgTrp  71
        |||||||||||||||||||||||||||||||||||||||||||||  ::::||
Db  13  CATGACCCGGCAGCTTCTCCTGGGCCTTGTCCTCTGGGCCAGTCGCCCCCGCCGCCAGTGG  72
```

```
Qy  72   LysGluArgAlaProSerSerSerAspThrAlaProGlyAlaMetProSerLeuSerVal  91
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  73   AAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCCCGGGGTGCAATGCCGAGCCTTCGGTA  132

Qy  92   ProAspArgArgGlyLeuLeuLeuAspProAlaAlaCysSerLysLysHisGlnProArg  111
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133  CCCGATCGCCGTGGAATGCCTGCTGGACCCTGCGGCCTGCTCCAAAACTCCACCAGCCCT  192

Qy  112  ValLeuHisCysHisValGlnAlaArgHisGlyCysProGlyProGlnLeuAlaLeuPro  131
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  GTCCTTCATTGCCACGTAACAGGCTCCGCATGGCCCGGGCCCCAACAGCTGGCCCTGCCT  252

Qy  132  AlaLeuAlaAspAlaAlaAsnValHisGlnLeuHisHisGlyLysCysProAlaValLeuHisGly  151
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253  GCAGCAGACGCCGCCAACGTCCACCAGCTGCACCATCACGGACGTGCAGCTGTTCTCCATGGC  312

Qy  152  SerLeuArgArgAlaGlnCysHisArgArgArgProProLeuGlyLeuGlnGlnLeuArgArgAla  171
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313  TCCCTACGTGCTCAATGTCACCGCCGTCCACCCCTGGGGCTCCAGCAGCAGCTTCGTGCC  372

Qy  172  PheHisAsnArgAlaIaHisHisGlnInAlaArgProSerArgArgArgGlaAlaProLysProPro  191
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  373  TTTCATAAACAGAGCACATCATCAAGCCCGACCCTCCAGAAGGCGTGCGCCTTAAGCCCCCT  432

Qy  192  Arg  192
         |||
Db  433  CGC  435
```

RESULT 6

```
LOCUS       AC005578      43297 bp    DNA    linear    PRI 01-SEP-1998
DEFINITION  Homo sapiens chromosome 19, cosmid F20887, complete sequence.
ACCESSION   AC005578
VERSION     AC005578.1  GI:3510237
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43297)
AUTHORS     Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andrelee,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE       Sequence analysis of a 2.5 Mb region in 19p13.3
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 43297)
AUTHORS     Lamerdin,J.E.
TITLE       Direct Submission
JOURNAL     Submitted (01-SEP-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Cosmid
            F20887 overlaps cosmid R33243 to the left from p telomere and
            overlaps BC84616 (CIT-B-144d21) to the right by at least 8 kb and
            43,297. Additional chromosome 19 map and sequence information may
            be obtained at: http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES    Location/Qualifiers
     source  1..43297
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="19"
             /map="19p13.3 between D19S883 and D19S325"
             /clone="F20887"
             /cell_line="UV5H19-5B"
             /clone_lib="LL19NC02 F chromosome 19-specific cosmid
             library"
             /note="Cosmid library constructed at LLNL from flow-sorted
             chromosomes from human-hamster hybrid UV5H19-5B, which
```

```
     CDS             carries chromosome 19 as its only human chromosome."
                     join(<177..283,457..644,1440..1477,2941..3048,5246..5374,
                     8654..8705,8927..9078,9226..9936,12267..12434,
                     16803..16868,17104..17181)
                     /note="Hypothetical human protein most similar to PID
                     9142314 and several ankyrin-like proteins"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F20887.1, partial CDS"
                     /protein_id="AAC33490.1"
                     /db_xref="GI:3510240"
                     /translation="SGATPLIIAAQMCHTDLCRLLLQQGRAAANDQDLQGRTALMLACE
                     GASPETVEVLLQGGAQPGITDALGQDAAHYGALAGDKLIHLLLQEAAQRPSPSALTE
                     DDSGEASSQDDRDAYEEIVRLRQERGRLLQKIRGLEQHKERRQQEROLSLHCCSQVQE
                     LQQLLVERQQEKESLGREVESLQSRLSLIENERENTSYDVTTLQDEEGPQGLTLRPPP
                     GAEVLLSROLSPSAQEHLASLQEQVAVLTRQNQELMEKVQILENFEKDETQMEVEALA
                     EVIPLAIYDSLRAEFDQLRRQHAEALQALRQQETREVPREEGAACGESEVVAGATATKN
                     GPTHMELNGSVAPETRKVNGAETIDEEAAGDETMEARTMEAEATGAEEATGAEATGAKVT
                     ETKPTGAEVREMETTEEEAANMETKPTGAQARDTETTGVEAMGVEATKTRAEEAAMQAY
                     GVGAGGAEPPVTGTTIMHEATSSRATCMESTCGVSATGVENPGVEATVQVREALFMKSERH
                     HLLYAIQGQMDEDVQRILSQILQMQRLQAQGR"
     repeat_region   318..346
                     /rpt_family="MIR"
     misc_feature    457..644
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 93.000"
     repeat_region   complement(1043..1258)
                     /rpt_family="AluSq"
     misc_feature    1440..1477
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 68.000"
     repeat_region   complement(1785..2083)
                     /rpt_family="AluSx"
     repeat_region   complement(2124..2421)
                     /rpt_family="AluSx"
     misc_feature    2734..2814
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 80.000"
     misc_feature    2941..3048
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 96.000"
     repeat_region   complement(3114..3264)
                     /rpt_family="LINE2"
     repeat_region   complement(3505..3808)
                     /rpt_family="AluSq"
     repeat_region   4033..4330
                     /rpt_family="AluSx"
     repeat_region   4495..4537
                     /rpt_family="LINE2"
     repeat_region   4550..4587
                     /rpt_family="LINE2"
     repeat_region   4595..4894
                     /rpt_family="AluY"
     repeat_region   4968..5001
                     /rpt_family="AluY"
     misc_feature    5246..5374
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 99.000"
     repeat_region   5527..5817
                     /rpt_family="AluSg"
     repeat_region   complement(5896..6085)
                     /rpt_family="(GGAA)n"
     repeat_region   complement(6086..6129)
                     /rpt_family="(GGAA)n"
     repeat_region   complement(6130..6323)
                     /rpt_family="AluSg/x"
     repeat_region   complement(6326..6608)
                     /rpt_family="AluSq"
     repeat_region   6700..6833
                     /rpt_family="AluSq"
     repeat_region   complement(6845..7093)
                     /rpt_family="AluJb"

     repeat_region   7249..7545
                     /rpt_family="AluSx"
     repeat_region   7791..7879
                     /rpt_family="L1"
     repeat_region   7880..8157
                     /rpt_family="AluSx"
     repeat_region   8161..8464
                     /rpt_family="AluJo"
     repeat_region   complement(8654..8705)
                     /rpt_family="AluSq"
     misc_feature    8654..8705
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 91.000"
     misc_feature    8960..9087
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 100.000"
     misc_feature    9226..10089
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 92.000"
     repeat_region   complement(10973..11275)
                     /rpt_family="AluSq"
     repeat_region   complement(11334..11425)
                     /rpt_family="(GGAA)n"
     repeat_region   complement(11426..11678)
                     /rpt_family="AluJo"
     repeat_region   11680..11981
                     /rpt_family="AluSx"
     repeat_region   11982..12179
                     /rpt_family="AluJb"
     misc_feature    12267..12434
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 100.000"
     repeat_region   complement(12536..12569)
                     /rpt_family="MIR"
     repeat_region   complement(12653..12945)
                     /rpt_family="AluSx"
     repeat_region   complement(12957..13135)
                     /rpt_family="MIR"
     repeat_region   complement(13229..13530)
                     /rpt_family="AluY"
     repeat_region   complement(13580..13751)
                     /rpt_family="AluJ"
     repeat_region   complement(13755..14043)
                     /rpt_family="AluSx"
     repeat_region   complement(14074..14203)
                     /rpt_family="AluJ"
     repeat_region   complement(14214..14318)
                     /rpt_family="MIR"
     repeat_region   14358..14430
                     /rpt_family="MIR"
     repeat_region   complement(14465..14667)
                     /rpt_family="MER20"
     repeat_region   complement(14697..14731)
                     /rpt_family="LINE2"
     repeat_region   14783..15083
                     /rpt_family="AluSx"
     repeat_region   15087..15137
                     /rpt_family="AluSx"
     repeat_region   15178..15228
                     /rpt_family="MER58B"
     repeat_region   complement(15624..15908)
                     /rpt_family="MER58B"
     repeat_region   complement(15919..16042)
                     /rpt_family="AluSg"
     repeat_region   16045..16077
                     /rpt_family="AluJo/FRAM"
     repeat_region   complement(16078..16246)
                     /rpt_family="(TA)n"
     repeat_region   complement(16251..16539)
                     /rpt_family="AluSx"
     repeat_region   complement(16540..16675)
                     /rpt_family="AluSc"
     misc_feature    16803..16868
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 89.000"
```

```
repeat_region    16896..16955
                 /rpt_family="MIR"
misc_feature     17104..17181
                 /note="predicted exon, program: grail2exons_human_1.3,
                 frame: 0, quality: excellent, score: 90.000"
repeat_region    complement(17628..17925)
                 /rpt_family="AluY"
repeat_region    complement(17926..18223)
                 /rpt_family="AluY"
repeat_region    18327..18613
                 /rpt_family="AluSx"
repeat_region    complement(18910..19211)
                 /rpt_family="AluJo"
```

US-09-921-994-2 (1-192) x AC005578 (1-43297)

```
Alignment Scores:
Pred. No.:
Score:                  9.75e-06         Length:         43297
                        374.50           Matches:        87
Percent Similarity:     61.64%           Conservative:   11
Best Local Similarity:  54.72%           Mismatches:     32
Query Match:            35.20%           Indels:         29
DB:                     9                Gaps:           8
```

```
Qy     53 AspProAlaAlaSerPro----------GlyProCysProLeuGlyGlnLeuProAlaLeu 69
          ::||||      :::             ||||
Db  25599 GAACCATCAGGGGCCACCGTGATGTTGGGGCCCATGCCCT------GCCCCCGG---     25646

Qy     70 ArgTrp-------LysGluArgAlaProSerSerArgAspThrAlaProGlyAla        85
          |||           |||
Db  25647 GGGTGGCACGGCCACCACCAGGACCCCACCCCGGATCCCCACCCGCTGCCCCTCCTGT     25706

Qy     86 --MetProSerLeuSerValProAsp----ArgArgGlyLeuLeuLeuAspProAlaLeuA 103
            :::|||      |||       ||||  :::|||
Db  25707 TCCTGCCCTCTCCTTGGGGTCCCGGGTCAGGCCCGGCAGTAGGAGGTGCTGAGGCCACAGG 25766

Qy    103 LeuCysSerLys----LeuHisGlnProArgValLeuHisCysHisValGlnAlaArgHisG 122
            :::          |||         ::|||
Db  25767 CACTCCCTCAGGGGCCTCAGCGAGCCCCAC------CCTGTGCAGGCTCGGCATG        25814

Qy    122 lyCysProGlyProGlnLeuLeuAlaLeuProLeuAlaLeuAlaAspAlaAsnValHisGlnLeuHisH 142
          |||||                              :::|||
Db  25815 GCTGCCCCGGGGCCCACAGCTGCCCCTGCCTGCAGCAGACGCCAACGTCCACCAGCTGCACC       25874

Qy    142 isHisGlyCysProAlaValLeuHisGlySerLeuArgAlaGlnCysHisArgArgProP  162
            |||||||    |||                   :::|||
Db  25875 ATCACGGAGAGTCACAGCTGTTCTCCATGGCTCCCTACGTGCTCAATGTCACCGCCGGTCCAC  25934

Qy    162 roLeuGlyLeuGlnLeuGlnLeuArgAlaPheHisAsnArgAlaHisHis---------    178
          ||||||||||||||
Db  25935 CCCTGGGGGCTCAGCAGCAGCTTCGTGCCTTTCATAACAGAGACACATCAGTGAGTGGGG   25994

Qy    179 -----------GlnAlaArgProSerArgArgArgAlaProLysPro                190
                     ||||||           |||
Db  25995 CGGCAGTGGGGGCGGCGGGGCGGGCGGGCTGCCGTCCCATCCCTTCCAGCTCCCCCACCCC  26047
```

RESULT 7

```
LOCUS       AX392987       445 bp    DNA      linear   PAT 23-MAR-2002
DEFINITION  Sequence 29 from Patent WO0214358.
ACCESSION   AX392987
VERSION     AX392987.1  GI:19701034
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Edmonds,B.T., Micanovic,R., Ou,W., Su,E.W., Tschang,S.H. and
            Wang,H.
TITLE       Novel secreted proteins and their uses
JOURNAL     Patent: WO 0214358-A 29 21-FEB-2002;
            ELI LILLY AND COMPANY (US)
FEATURES             Location/Qualifiers
```

```
source          1..445
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
CDS             25..333
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAD29041.1"
                /db_xref="GI:19701035"
                /translation="MIPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAV
                DCSWTLPPAPNSTSPVSFIATYRSESLEGGLRDTGLPGEPRTLALRALGSGKLETPTS"
mat_peptide     76..330
                /product="unnamed"
BASE COUNT     124 a    138 c    115 g    68 t
ORIGIN
```

US-09-921-994-2 (1-192) x AX392987 (1-445)

```
Alignment Scores:
Pred. No.:
Score:                  2.18e-07         Length:         445
                        373.00           Matches:        69
Percent Similarity:     92.11%           Conservative:   1
Best Local Similarity:  90.79%           Mismatches:     6
Query Match:            35.06%           Indels:         0
DB:                                      Gaps:           0
```

```
Qy     52 HisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArgTrp   71
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     24 CATGACCCCGCAGCTTCTCCTGGCCCTGCCCTTGTCCTCTGGGCCCCTGCCCGCCCTGTGG  83

Qy     72 LysGluArgAlaProSerSerArgAspThrAlaProGlyAlaMetProSerLeuSerVal   91
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     84 AAGGAAAGGCCCCCACCAGCTCTGCACACTGCCCGGGGTGCAAATGCCGAGCCTCTCGGTA 143

Qy     92 ProAspArgArgGlyLeuLeuLeuAspProAlaLeuAlaCysSerLysLeuHisGlnProArg 111
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    144 CCCGATGCGCCGGGGATTGGCTCCTGGACCCTGCCGCGCCTCCCAAAACTCACCAGCCCCGT 203

Qy    112 ValLeuHisCysHisValGlnAlaArgHisGlyCysProGlyGln                 127
          |||||||||||||||||||||
Db    204 GTCCTTCATTGCCACGTACAGGTTCGGGAGAGCCTGGAAGGGGGGCCCTCAG          251
```

RESULT 8

```
LOCUS       AC027272/c
                         179607 bp    DNA      linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 19 clone RP11-486P13, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC027272
VERSION     AC027272.4  GI:9838224
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 179607)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 179607)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (28-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:8570236.
            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0486P13
            -------------- Summary Statistics --------------
```

```
Sequencing vector: M13; 95%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 95% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 170867 bases at least Q40
Consensus quality: 174124 bases at least Q30
Consensus quality: 176036 bases at least Q20
Consensus quality: 188000; agarose-fp
Insert size: 188000; agarose-fp
Insert size: 179350; sum-of-contigs
Quality coverage: 4.55 in Q20 bases; agarose-fp
Quality coverage: 4.80 in Q20 bases; sum-of-contigs

*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 18 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*       1        1338: contig of 1338 bp in length
*    1339        1438: gap of unknown length
*    1439        3330: contig of 1892 bp in length
*    3331        3430: gap of unknown length
*    3431        5429: contig of 1999 bp in length
*    5430        5529: gap of unknown length
*    5530        8107: contig of 2578 bp in length
*    8108        8207: gap of unknown length
*    8208       11180: contig of 2973 bp in length
*   11181       11280: gap of unknown length
*   11281       14039: contig of 2759 bp in length
*   14040       14139: gap of unknown length
*   14140       17230: contig of 3091 bp in length
*   17231       17330: gap of unknown length
*   17331       21334: contig of 4004 bp in length
*   21335       21434: gap of unknown length
*   21435       26877: contig of 5443 bp in length
*   26878       26977: gap of unknown length
*   26978       31453: contig of 4476 bp in length
*   31454       31553: gap of unknown length
*   31554       36515: contig of 4962 bp in length
*   36516       36615: gap of unknown length
*   36616       46678: contig of 10063 bp in length
*   46679       46778: gap of unknown length
*   46779       57999: contig of 11221 bp in length
*   58000       58099: gap of unknown length
*   58100       74742: contig of 16643 bp in length
*   74743       74842: gap of unknown length
*   74843       92504: contig of 17662 bp in length
*   92505       92604: gap of unknown length
*   92605      119545: contig of 26941 bp in length
*  119546      119645: gap of unknown length
*  119646      147254: contig of 27609 bp in length
*  147255      147354: gap of unknown length
*  147355      179607: contig of 32253 bp in length.

FEATURES             Location/Qualifiers
     source          1..179607
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-486P13"
     misc_feature    1..1338
                     /note="assembly_name:Contig8"
     misc_feature    1439..3330
                     /note="assembly_name:Contig9"
     misc_feature    3431..5429
                     /note="assembly_name:Contig10"
     misc_feature    5530..8107
                     /note="assembly_name:Contig11"
     misc_feature    8208..11180
                     /note="assembly_name:Contig12"
     misc_feature    11281..14039
```

```
     misc_feature    14140..17230
                     /note="assembly_name:Contig13"
     misc_feature    17331..21334
                     /note="assembly_name:Contig14"
                     /clone_end:SP6
                     vector_side:left"
                     21435..26877
                     /note="assembly_name:Contig15
     misc_feature    26978..31453
                     /note="assembly_name:Contig16"
     misc_feature    31554..36515
                     /note="assembly_name:Contig17"
     misc_feature    36616..46678
                     /note="assembly_name:Contig18"
     misc_feature    46779..57999
                     /note="assembly_name:Contig19"
     misc_feature    58100..74742
                     /note="assembly_name:Contig20"
     misc_feature    74843..92504
                     /note="assembly_name:Contig21"
     misc_feature    92605..119545
                     /note="assembly_name:Contig22"
     misc_feature    119646..147254
                     /note="assembly_name:Contig23"
     misc_feature    147355..179607
                     /note="assembly_name:Contig24"
     misc_feature    147355..179607
                     /note="assembly_name:Contig25"

BASE COUNT    41107 a   48085 c   48884 g   39812 t   1719 others
ORIGIN
```

```
US-09-921-994-2 (1-192) x AC027272 (1-179607)

Alignment Scores:
Pred. No.:
Score:                      0.000257        Length:            179607
Score:                    351.50            Matches:               89
Percent Similarity:        51.05%           Conservative:           8
Best Local Similarity:     46.84%           Mismatches:            35
Query Match:               33.04%           Indels:                60
DB:                         2                Gaps:                   7
```

```
Qy       12  TrpAlaSerCysPro---ProArgSerGlyArgLysGlyProProAlaAlaLeuThrLeu  30
                |||                |||||||      ||||||
Db  158705  TGGGGCGCCCATGCCCTTGCCCCCGGGGGGGTGGCGCACGGCCACACCAGG------    158659

Qy       31  ProArgValGlnCysArgAlaSerArgTyrProIleAlaValAlaSpCysSerTrpThrLeu  50
                                |||||||                    |||
Db  158658  -----------------------------------------ACCCCA    158653

Qy       51  ProHisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArg  70
                |||  ||||||       ||                 |||
Db  158652  CCCCGGGGATCCCCATCCGCTGCCCCCTCCT------------GTTCCTGCCTCTCCT    158608

Qy       71  TrpLysGluArgAlaProSerSerAspThrAlaProGlyAlaMetProSerLeuSer  90
                                            |||||||      |||::::
Db  158607  TGG-----------------------------GGTCCCGGGTCAGGCCCGGCAGTAGGA    158578

Qy       91  ValProAspArgArgGly-LeuLeuLeuAspProAlaAlaCysSerLysLysHisGlnPr  110
                :::              |||    |||                 |||
Db  158577  GGTGCTGAGGGCCACAGGCACTC-----------CCTGAGGGCGCCTCACGGAGCCGCCACCCT--    158529

Qy      110  oArgValLeuHisCysHisSerValGlnLLnAlaArgHisGlyCysHisProGlyProGlnLeuAlaLe  130
                |||||              |||||||||||||||:||::||||||||||||||||||
Db  158528  --------------------GTGCAGGCTCGGCATGGCTGCCCGCGGGCCAGAGTTGGC-CT    158489

Qy      130  uProAlaIaAlaAspAlaAsnValHisGlnLeuHisHisHisSglyCysProAlaValLeuuHi  150
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db  158488  GCCTCGACAGACAGACGCCAACGTCCACCAGCTGCACCATAGAGTCCAGCTCGTTCTCCA    158429

Qy      150  sGlySerLeuArgAlaGlnCysHisSerArgArgProProLeuGlyLeuGlnLeuGlnLeuAr  170
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  158428  TGGCTCCCTACGTGCTCAAGTCACCGCGTCACCCCTGGGGCGTCCAGCAGCAGCCTTCG    158369

Qy      170  gAlaPheHisAsnArgAlaHisHisHis------------------------GlnAlaAr  181
```

Alignment Scores:
Pred. No.:

```
Db  158368 gProSerArgArgArgAlaProLysPro 190
Qy     181 gProSerArgArgArgAlaProLysPro 190
           ||||||||||||||||||||||||||||
Qy     181 TGCCTTTCATTAACAGACAGAGACACATCAGTGAGTGGGTGGGGTGGGGCGGCGAGTGGGGCGCCGGGGGGGGCT 158309
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||
Db  158308 GCCGTCTTCCTTCCAGCTCCCCACCC 158281
```

RESULT 9
BC008209
LOCUS   BC008209   1131 bp   mRNA   linear   ROD 07-AUG-2002
DEFINITION   Mus musculus, Epstein-Barr virus induced gene 3, clone MGC:11457
                    IMAGE:3154957, mRNA, complete cds.
ACCESSION   BC008209
VERSION   BC008209.1 GI:14198294
KEYWORDS   MGC.
SOURCE   house mouse.
  ORGANISM   Mus musculus
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1131)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-MAY-2001) National Institutes of Health, Mammalian
                    Gene Collection (MGC), Cancer Genomics Office, National Cancer
                    Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                    USA
COMMENT   NIH-MGC Project URL: http://mgc.nci.nih.gov
                    Contact: MGC help desk
                    Email: cgapbs-r@mail.nih.gov
                    Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
                    cDNA Library Preparation: Life Technologies, Inc.
                    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                    DNA Sequencing by: Baylor College of Medicine Human Genome
                    Sequencing Center
                    Center code: BCM-HGSC
                    Web site: http://www.hgsc.bcm.tmc.edu/cdna/
                    Contact: amg@bcm.tmc.edu
                    Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
                    Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
                    Richards, S., Gibbs, R.A.
                    clone distribution: MGC clone distribution information can be found
                    through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                    Series: IRAK Plate: 16 Row: c Column: 19
                    This clone was selected for full length sequencing because it
                    passed the following selection criteria: matched mRNA gi: 7657051.
REMARK
FEATURES   Location/Qualifiers
     source   1..1131
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="129,C57BL/6J,FVB/N"
                     /clone="MGC:11457 IMAGE:3154957"
                     /tissue_type="Mammary tumor. Brca1-/fl; MMTV-Cre model. 10
                     months old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam3"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   188..874
                     /codon_start=1
                     /product="Epstein-Barr virus induced gene 3"
                     /protein_id="AAH08209.1"
                     /db_xref="GI:14198295"
                     /db_xref="LocusID:50498"
                     /translation="MSKLLFLSLALWASRSPGYTETALVALSQPRVQCHASRXPVAVD
                     CSWPLQAENSTRTSFIATYRLGVATOQOSOPCLQRSPQASRCTIPDVHLFSTVPYM
                     LNTAVHPGGASSSLLAFVAERIIKPDPPEGVRLRTAGORLQVIWHPPASWPEPDIFS
                     LKYRLRYRRGASHFRQVGPIEATTFTLRNSKPHAKYCIQVSAQDLIDYGKPSDWSLP
                     GQVESAPHKP"
BASE COUNT   242 a   374 c   268 g   247 t
ORIGIN
```

242 a   0.0017   Length: 1131

US-09-921-994-2 (1-192) x BC008209 (1-1131)

```
                   Score:                 281.50          Matches:            79
                   Percent Similarity:    46.94%          Conservative:       13
                   Best Local Similarity: 40.31%          Mismatches:         69
                   Query Match:           26.46%          Indels:             36
                   DB:                                    Gaps:                6

Qy      14 SerCysPro----------------------ProArgSerGlyArgLysGlyProProAlaAla 27
                                                 ||||               ||||   |||
Db      50 TCCTGCCCCCATACACTGGACAACTGAGCCCACACTG-GGCAGGTTCCTTCCCTGGGGCC--- 105

Qy      28 LeuThrLeuProArgValGlnCysArgAlaSerArgTyrProIleAlaValAspCysSer 47
           |||
Db     193 CAAGCTGCTTCGTTCCTGTTCACTTGCCCTCTGGGCCAG------CCGCTCCCCTGGGTTCAC 246

Qy      74 ---ArgAlaProSerSerSerAspThrAlaProGlyIleAlaMetProSerLeuSerValPro 92
                  :::                                 ||   :::    |||
Db      74 TGAAACACAGCTCTCGTGGGTCTTAAGCCACGCCCCAGAGTGCAATGCCATTGCTTCTCGGTATCC 306

Qy     106 ----------------TrpThrLeuProHisAspPro---------------------------- 54
                           |||                                     |||
Db     247 TGAAACACAGCTCTCGTGGGTCTTAAGCCACGCCCCAGAGTGCAATGCCATTGCTTCTCGGTATCC 306

Qy     133 TCACCCTCTCTCTGATGGGTCACTAACTCGGATCCAGGAACAGAGCCACCAGAGCAGTCTC 192
           |||||||||
Db     133 TCACCCTCTCTCTGATGGGTCACTAACTCGGATCCAGGAACAGAGCCACCAGAGCAGTGTC 192

Qy      55 --AlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArgTrpLysGlu 73
                  ||||||||||||||||||||||     |||
Db     307 CGTGGCCGGTGGACTCGTTCCTGGAACTCCTCCAACTCCACCAGATCCACGTC 366

Qy      93 AspArgArgGlyLeuLeuLeuAspProAlaCysSerLysSerLeuHisGlnProArgVal 112
                  :::                    |||||||||||                  :::
Db     427 ACGGAGCGCCCCAGGCCTCCCCGATGCACCATCCCCGACCGTGTTCTCCACGGTGCC 486

Qy     113 LeuHisCysHisLeuSValGlnAlaArgHisGlyCysProGlyProGlnLeuLeuAlaLeuProAla 132
           |||||||||               |||||||          :::     ||||||||||
Db     367 CTTCATTGCCACTTACAGGCTCGGTGTGGCCACCCAGCACGGGCACGCCCTGCCTACA 426

Qy     133 AlaAspAlaAsnValHisGlnLeuHisHisHisGlyCysProLeuValLeuHisGlySer 152
                  :::                    ||                    :::
Db     487 CTACATCGCTAAAATGTCACTGCAGTGCACCATCCCCGACCGTGTTCTCCACGGTGCC 546

Qy     153 LeuArgAlaGlnCysHisHisGlnAlaArgProSerArgArgArgAlaPro 188
                  |||                ||                   |||
Db     427 ACGGAGCGCCCCAGGCCTCCCCGATGCACCATCCCCGACCGTGTTCTCCACGGTGCC 486

Qy     173 HisAsnArgAlaHisHisGlnAlaArgProSerArgArgArgAlaPro 172
                  |||||||||||||||       |||||||||||
Db     487 CTACATCGCTAAAATGTCACTGCAGTGCACCATCCCCGACCGTGTTCTCCACCCTTT 546

Qy     173 HisAsnArgAlaHisHisGlnAlaArgProSerArgArgArgAlaPro 188
                  |||||||||||
Db     547 TGTGGCCTGAGCGGAATCATCAAGCCACGACCCTCCGGAAGGCGTGCGCCT 594
```

RESULT 10
AF013114
LOCUS   AF013114   1161 bp   mRNA   linear   ROD 21-AUG-1997
DEFINITION   Mus musculus cytokine receptor-like molecule (EBI3) mRNA, complete
                    cds.
ACCESSION   AF013114
VERSION   AF013114.1 GI:2338440
KEYWORDS   .
SOURCE   Mus musculus
  ORGANISM   Mus musculus
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1161)
  AUTHORS   Nomura,H., Yaguchi,N., Hanyuu,C., Maeda,M., Kikuchi,Y., Nakata,Y.,
                    Kojima,T., Tulin,E.E. and Hasegawa,M.
  TITLE   Mouse homolog of human Epstein-Barr virus-induced gene 3 (EBI3)
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1161)
  AUTHORS   Nomura,H., Yaguchi,N., Hanyuu,C., Maeda,M., Kikuchi,Y., Nakata,Y.,
                    Kojima,T., Tulin,E.E. and Hasegawa,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-JUL-1997) Cytokine Research Program, CHUGAI Research
                    Institute for Molecular Medicine, Inc., 153-2, Nagai, Niihari,
                    Ibaraki 300-41, Japan
FEATURES   Location/Qualifiers

```
source          1..1161
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone_lib="Clontech Marathon-Ready library #7459-1"
                /dev_stage="15-day embryo"
gene            1..1161
                /gene="EBI3"
                /note="MMEBI3; MUSEBI3"
CDS             223..909
                /gene="EBI3"
                /note="similar to human EBI3 protein"
                /codon_start=1
                /product="cytokine receptor-like molecule"
                /protein_id="AAB67115.1"
                /db_xref="GI:2338441"
                /translation="MSKLLFLSLALWASRSPGYTETALVALSQPRVQCHASRYPVAVD
                CSWTPLQAPNSTRSTFSPIATYRLGVATQOQOSOPCLQRSPQASRCTIPDVHLFSTVPYM
                LNVTAVHPGGASSSLLAFVAERIIKPDPPEGVRLRTAGQRLQVLHPPASWPFPDIFS
                LKYRLRYRRGASHFRQVGPIEATTFTLRNSKPHAKYCIQVSAQDLTDYGKPSDWSLP
                GOVESAPHKP"
BASE COUNT       390 a     272 g     255 t     244 c
ORIGIN

US-09-921-994-2 (1-192) x AF013114 (1-1161)

Alignment Scores:

Pred. No.:             0.00173
Score:                 281.50          Length:         1161
Percent Similarity:    46.94%          Matches:        79
Best Local Similarity: 40.31%          Conservative:   13
Query Match:           26.46%          Mismatches:     69
DB:                    10              Indels:         36
                                       Gaps:           6

Qy    14 SerCysPro---------------------ProArgSerGlyArgLysGlyProProAlaAla 27
         ||||||||||                        |||||||||||||||||||||
Db    85 TCCTGCCCATACACTGGACAACTGAGCCACACTG-GGCAGGTCCTTCCCTGGGCGCC--- 140

Qy    28 LeuThrLeuProArgValGlnCysArgAlaSerArgTyrProIleAlaValAspCysSer 47
         |||||||||||||||||::::||:::||:|
Db   141 --------------AGGTTCCCTGTGTGAGTTCCCCTGTCCT 167

Qy    48 ----------------TrpThrLeuProHisAspPro------------------- 54
              |||                      |||
Db   168 TCACCCTCTCTGATGGGTCACTAACTCGGATCCAAGGAACAGAGACCACAGAGCANTGTC 227

Qy    55 ---AlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuArgTrpLysGlu 73
            |||||  |||||||||||||||||||||||||||||||||||||
Db   228 CAAGCTGCTCTTCCGTCACTTGCCCTCTGGGCCAG------CCGCTCCCTGGTTACAC 281

Qy    74 ---ArgAlaProSerSerAspThrAlaProGlyAlaMetProSerLeuSerValPro 92
            |||       :::
Db   282 TGAAACAGCTCTCGTGGCTCTTAAGCCAGCCCCAGAGTGCAANTGCCATTGCCTTCTCGGTATCC 341

Qy    93 AspArgArgArgGlyLeuLeuLeuAspProAlaAlaCysSerLysLeuSerLeuHisGlnProArgVal 112
         |||||||||||||||::       :::
Db   342 CGTGCCGTGGACTCCTCCTGGACTCCTCCCAACTCACCAGATCCAGTC 401

Qy   113 LeuHisCysHisValGlnAlaAlaArgHisHisGlyCysProGlyProGlnLeuAlaLeuProAla 132
         |||:::||||::||::|||||
Db   402 CTTCANTTGCCACTTACAGCGTCGGTGTGGCACCCAGCAGACCCTGCCTACA 461

Qy   133 AlaAspProAlaAsnValHisSerLeuHisHisHisGlyLeuHisGlnLeuAlaLeuValLeuHisGlySer 152
         |||:::|||::|||||
Db   462 ACGGAGCCCCAGCGCCTCCCGATGCACCATCCCCGACGTCCACCTGTTCTCCACGGTGCC 521

Qy   153 LeuArgLeuAlaGlnCysHisSerArgArgProProLeuGlyLeuGlnLeunLeuArgGlnAlaPhe 172
         |||::||||||||||||:::
Db   522 CTACATGCTAAATGTCACTGCAGTGCACCCAGGGGGCGGCGCCAGCAGCAGCGCCTCGTAGCCTT 581

Qy   173 HisAsnArgAlaHisHisGlnLnAlaArgProSerArgArgArgAlaPro 188
         |||::||||||:::
Db   582 TGTGGCTGACGAATCATCAAGCCGGAACCCTCCGAAGGCGTGCGCCT 629
```

```
RESULT 11
AC073737
AC073737/c
LOCUS       AC073737   193355 bp   DNA   linear   HTG 18-JUL-2000
DEFINITION  Mus musculus clone RP23-264G1, WORKING DRAFT SEQUENCE, 24 ordered
            pieces.
ACCESSION   AC073737
VERSION     AC073737.2  GI:9256775
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Mus musculus.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 193355)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Mouse
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 193355)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:8810354.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 1838393
            Center clone name: RPCI-23_264G1
            -----
            Summary Statistics
            Consensus quality: 175100 bases at least Q40
            Consensus quality: 187679 bases at least Q30
            Consensus quality: 189929 bases at least Q20
            Estimated insert size: 87000; agarose-fp estimation
            Estimated insert size: 192255; sum-of-contigs estimation
            Quality coverage: 10.37 in Q20 bases: agarose-fp estimation
            Quality coverage: 4.69 in Q20 bases: sum-of-contigs estimation
            NOTE: This is a 'working draft' sequence. It currently
            consists of 24 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces
            is believed to be correct as given, however the sizes
            of the gaps between them are based on estimates that have
            provided by the submitter.
            This sequence will be replaced
            by the finished sequence as soon as it is available and
            the accession number will be preserved.
     *             1:    10615: contig of 10615 bp in length
     *         10616:    10715: gap of unknown length
     *         10716:    40715: contig of 30013 bp in length
     *         40716:    40728: gap of unknown length
     *         40729:    40828: gap of unknown length
     *         40829:    43385: contig of 2557 bp in length
     *         43386:    43485: gap of unknown length
     *         43486:    60463: contig of 16978 bp in length
     *         60464:    60563: gap of unknown length
     *         60564:    63757: contig of 3194 bp in length
     *         63758:    63857: gap of unknown length
     *         63858:    72484: contig of 8627 bp in length
     *         72485:    72584: gap of unknown length
     *         72585:    80251: contig of 7667 bp in length
     *         80252:    80351: gap of unknown length
     *         80352:   101558: contig of 21207 bp in length
     *        101559:   101658: gap of unknown length
     *        101659:   113941: contig of 12283 bp in length
     *        113942:   114041: gap of unknown length
     *        114042:   118302: contig of 4261 bp in length
     *        118303:   118402: gap of unknown length
     *        118403:   122141: contig of 3739 bp in length
     *        122142:   122242: gap of unknown length
     *        122242:   141898: contig of 19657 bp in length
     *        141899:   141998: gap of unknown length
     *        141999:   145605: contig of 3607 bp in length
```

```
US-09-921-994-2 (1-192) x AC073737 (1-193355)

DB:

Query Match:
Best Local Similarity:   18.19%
Percent Similarity:      32.02%
Score:                   40.35%
                         193.50
Pred. No.:               360
Alignment Scores:

                         Length:        193355
                         Matches:       73
                         Conservative:  19
                         Mismatches:    74
                         Indels:        62
                         Gaps:          8

ORIGIN
BASE COUNT   49405 a  44453 c  45086 g  52104 t  2307 others

FEATURES             Location/Qualifiers
     source          1..193355
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-264G1"
                     /clone_lib="RPCI mouse BAC library 23"

*  145506
   145705:  gap of unknown length
*  148556
   148855:  contig of 2850 bp in length
*  148656
   151425:  contig of 2270 bp in length
*  151426
   151525:  gap of unknown length
*  151526
   158286:  contig of 6761 bp in length
*  158287
   158386:  gap of unknown length
*  158387
   160637:  contig of 2252 bp in length
*  160638
   160738:  gap of unknown length
*  160739
   173388:  contig of 12650 bp in length
*  173389
   173488:  gap of unknown length
*  173489
   179077:  contig of 5589 bp in length
*  179078
   179177:  gap of unknown length
*  179178
   184518:  contig of 5341 bp in length
*  184519
   184618:  gap of unknown length
*  184619
   189175:  contig of 4457 bp in length
*  189176
   189275:  gap of unknown length
*  189276
   191154:  contig of 1879 bp in length
*  191155
   191254:  gap of unknown length
*  191255
   192103:  contig of 849 bp in length
*  192104
   192203:  gap of unknown length
*  192204
   193355:  contig of 1152 bp in length
```

```
Qy    24 ProProAla------------AlaLeuThrLeuProArgValGlnCysArgAlaSerArgTyrPro  41
            ||||||||            ::|  |||||||||||||||||:|||
Db112289 CCTCCAGCTGCTCTCTAAGCCAGCCCAGAGTGCAATGCCACTGCTTCTCTGGTATCCC 112230

Qy    42 IleAlaValAlaAspCysSerTrpThr-LeuProHisAspProAlaAlaSerProGlyProCys  61
            |||||||||||||||||||||| |||               |||  |||||||||||||
Db112229 GTGCGCGTGGACTGCTCCTGGACTCCCTCCAGGCTCCCA--ACTCCACCAGATCCACG 112173

Qy    61 sProLeu-GlyGlnLeuProAlaLeuuArgTrpLys------GluArgAlaProSer-Ser  78
            |||     :::||||||||  |||||||||||          |||  ||||||| |||
Db112172 TCCTTCATTGCCACTTACAGGTCAGAAAATGGAAAGGGTTTAGGAGGAAGCCAAGTATCA 112113

Qy    79 SerAspThr---------------AlaProGlyAlaMetProSerLeu  89
            |||                     ::|||||||||||||||||||
Db112112 AGAGATTCCAAATGTCCTGATCCACTCAGGTAGGAGCAGCCTAGGCCCTTGGGAACCTG 112053

Qy    90 SerValProAspArg---------------GlyLeuLeuAspPro 101
            |||                          :::|||  :::|||
Db112052 CTTCTAAAGCAAACAGAACAAGTGTGGGTTGTGGTATTCTAGGGAGTTCTTCAGAATCCA 111993

Qy   102 AlaAlaCysSer------------ 105
            |||
Db111992 CAGTCCCAAGTCCCCCTGAGCATATCACCCCTGTGAGCGTGACCCACAGTCAGCTGTGG 111933

Qy   106 ------LysLeuHisGlnProArgValLeuHis------------CysHisVal 117
                  |||                                      ::|
Db111932 TGAGCGTGGGGCCTCACGGGGCACCAGGTTCCTTCGAGATTCACCATGGTGCTTCCCATG 111873

Qy   118 GlnAlaAlaArgHisGlyCysProGlyProGlnLeuAlaLeuProAlaLeuAAspAlaAsnVal 137
            |||||||||| |||  |||                         :::  :::
Db111872 CAGGCTCGGTTGTGGCCACCCAGCAGCAGCCCTACAACGAGCGCCCCAGGC 111813

Qy   138 HisGlnLeuHisHisHisGlyCysProAlaValLeuHisGlySerLeuArgLeuAlaGlnCys 157
            ::||||||||| |||||||||||||||||||:||:|||||||||:::|||  :::|||
Db111812 CTCCGAGATCCACCAATCCCCGACGTGCACGTGCCACCGTGTTCTCCACGGTGCCCTACACTGCTAAATGT 111753

Qy   158 HisArgArgProProLeuGlnLeuGlnGlnLeuArgAlaProAlaPheHisAsnArgAlaHis 177
            |||  |||                                           :::
Db111752 CACTGCAGTGCACCCAGGCGGCCGGCGGCCAGCAGCACGCCTCTAGCCGCTTTGGCTGGAGCGAAT 111693

Qy   178 ----------------------HisGlnAlaArgProSerArg 184
                                  |||
Db111692 CAGTGAGTGGTGCTGTCCTGCCGTCCCCACTTGCCTCCATCCCGTGCCTCCTCCCTGC 111633

Qy   185 ArgArgAlaProGlyLysPro 190
            ::|||  :::|||  |||
Db111632 CGCCGCCCCCCCCCTACCCCC 111615
```

RESULT 12
AC118074/c

```
LOCUS       AC118074              176359 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-47J17, *** SEQUENCING IN PROGRESS
            ***, 69 unordered pieces.
ACCESSION   AC118074
VERSION     AC118074.4  GI:21746302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
```

```
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (13-APR-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 176359)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 14, 2002 this sequence version replaced gi:20177536.
COMMENT   
          ------------------------- Genome Center
          
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          ------------------------- Project Information
          
          Center project name: GUEB
          Center clone name: CH230-47J17
          ------------------------- Summary Statistics
          
          Sequencing vector: Plasmid;
          Chemistry: Dye-terminator Big Dye: 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 115372 bases at least Q40
          Consensus quality: 121114 bases at least Q30
          Consensus quality: 124422 bases at least Q20
          
          NOTE: Estimated insert size may differ from sequence length
          (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          NOTE: This is a 'working draft' sequence. It currently
          consists of 69 contigs. The true order of the pieces
          is not known and their order in this sequence record is
          arbitrary. Gaps between the contigs are represented as
          runs of N, but the exact sizes of the gaps are unknown.
          This record will be updated with the finished sequence
          as soon as it is available and the accession number will
          be preserved.

          *        1     1219: contig of 1219 bp in length
          *     1220     1319: gap of unknown length
          *     1320     2744: contig of 1425 bp in length
          *     2745     2844: gap of unknown length
          *     2845     4066: contig of 1222 bp in length
          *     4067     4166: gap of unknown length
          *     4167     5794: contig of 1628 bp in length
          *     5795     5894: gap of unknown length
          *     5895     6936: contig of 1042 bp in length
          *     6937     7036: gap of unknown length
          *     7037     8766: contig of 1731 bp in length
          *     8767     8866: gap of unknown length
          *     8867    10546: contig of 1679 bp in length
          *    10547    10646: gap of unknown length
          *    10647    12271: contig of 1625 bp in length
          *    12272    12371: gap of unknown length
          *    12372    13483: contig of 1112 bp in length
          *    13484    13583: gap of unknown length
          *    13584    14951: contig of 1368 bp in length
          *    14952    15051: gap of unknown length
          *    15052    16270: contig of 1219 bp in length
          *    16271    16370: gap of unknown length
          *    16371    17804: contig of 1434 bp in length
          *    17805    17904: gap of unknown length
          *    17905    19394: contig of 1490 bp in length
          *    19395    19494: gap of unknown length
          *    19495    20712: contig of 1218 bp in length
          *    20713    20812: gap of unknown length
          *    20813    22799: contig of 1987 bp in length
          *    22800    22899: gap of unknown length
          *    22900    24177: contig of 1278 bp in length
          *    24178    24277: gap of unknown length
          *    24278    25821: contig of 1544 bp in length
          *    25822    25921: gap of unknown length
          *    25922    27784: contig of 1863 bp in length

          *    27785    27884: gap of unknown length
          *    27885    28934: contig of 1050 bp in length
          *    28935    29034: gap of unknown length
          *    29035    30354: contig of 1320 bp in length
          *    30355    30454: gap of unknown length
          *    30455    31701: contig of 1247 bp in length
          *    31702    31801: gap of unknown length
          *    31802    33337: contig of 1536 bp in length
          *    33338    33437: gap of unknown length
          *    33438    34728: contig of 1291 bp in length
          *    34729    34828: gap of unknown length
          *    34829    36591: contig of 1763 bp in length
          *    36592    36691: gap of unknown length
          *    36692    38003: contig of 1312 bp in length
          *    38004    38103: gap of unknown length
          *    38104    40127: contig of 2024 bp in length
          *    40128    40227: gap of unknown length
          *    40228    41691: contig of 1464 bp in length
          *    41692    41791: gap of unknown length
          *    41792    43535: contig of 1744 bp in length
          *    43536    43635: gap of unknown length
          *    43636    45466: contig of 1831 bp in length
          *    45467    45566: gap of unknown length
          *    45567    47021: contig of 1455 bp in length
          *    47022    47121: gap of unknown length
          *    47122    48801: contig of 1680 bp in length
          *    48802    48901: gap of unknown length
          *    48902    49935: contig of 1034 bp in length
          *    49936    50035: gap of unknown length
          *    50036    51484: contig of 1449 bp in length
          *    51485    51584: gap of unknown length
          *    51585    53366: contig of 1782 bp in length
          *    53367    53466: gap of unknown length
          *    53466    55706: contig of 2240 bp in length
          *    55707    55806: gap of unknown length
          *    55807    57484: contig of 1678 bp in length
          *    57485    57584: gap of unknown length
          *    57585    59524: contig of 1940 bp in length
          *    59525    59624: gap of unknown length
          *    59625    61504: contig of 1880 bp in length
          *    61505    61604: gap of unknown length
          *    61605    63038: contig of 1434 bp in length
          *    63039    63138: gap of unknown length
          *    63139    65148: contig of 2010 bp in length
          *    65149    65248: gap of unknown length
          *    65249    68165: contig of 2917 bp in length
          *    68166    68265: gap of unknown length
          *    68266    70437: contig of 2172 bp in length
          *    70438    70537: gap of unknown length
          *    70538    72778: contig of 2241 bp in length
          *    72779    72878: gap of unknown length
          *    72879    74865: contig of 1987 bp in length
          *    74866    74965: gap of unknown length
          *    74966    78152: contig of 3187 bp in length
          *    78153    78252: gap of unknown length
          *    78253    80830: contig of 2578 bp in length
          *    80831    80930: gap of unknown length
          *    80931    83775: contig of 2845 bp in length
          *    83776    83875: gap of unknown length
          *    83876    86764: contig of 2889 bp in length
          *    86765    86864: gap of unknown length
          *    86865    90526: contig of 3662 bp in length
          *    90527    90626: gap of unknown length
          *    90627    92648: contig of 2022 bp in length
          *    92649    92748: gap of unknown length
          *    92749    96414: contig of 3666 bp in length
          *    96415    96514: gap of unknown length
          *    96515    99630: contig of 3116 bp in length
          *    99631    99730: gap of unknown length
          *    99731   102795: contig of 3065 bp in length
          *   102796   102895: gap of unknown length
          *   102896   105478: contig of 2583 bp in length
```

```
Alignment Scores:
Pred. No.:
Score:                  1.98e+03        Length:        176359
Percent Similarity:     173.50          Matches:       74
Percent Similarity:     38.89%          Conservative:  17
Best Local Similarity:  31.62%          Mismatches:    65
Query Match:            16.31%          Indels:        79
                        2               Gaps:          9

US-09-921-994-2 (1-192) x AC118074 (1-176359)

Qy    28  LeuThrLeuProArgValGlnCysLysArgAlaSerArgTyrProIleAlaValAspCysSer  47
          |||:::                                                    |||
Db 67089  CTGACCCAGCCCAGAGTGCCATGCCATGCTTCTCGGTACCCTGTGGGTCGTGGGACTGCGTCC 67030

Qy    48  TrpThrLeuProHisAspProAlaAlaSer--ProGlyProCysProLeuGlyGlnLeuPr  67
          :::|||                         |||||||||||||
Db 67029  TGG--ATGCCACTCCAGGTCCCAACTCCACCAGAGCCATGTCCTTCATTGCCACTTAC 66973

Qy    67  oAlaLeuArg-TrpLysGlu-------------------ArgAlaPro-  76
          |||                                    |||
Db 66972  AGGTGAGAGAAATGGAAAGGGTTTGGATGAGGCTGAGAATCAAGAGGATTCCAAATGTCCTG 66913

Qy    76  ------------------SerSerAspThrAlaProGlyAlaMetPr  87
                            |||||||||||
Db 66912  ATCCACTCAGGGTAGGAGGCAGCCTAGGCCCCATGGGAGCCTGCTTCTAAAGCAACAGAA 66853

Qy    87  ro-SerLeuSerValProAspArgArgGlyLeuLeuLeuAspPro---AlaAlaCysSer 105
          |||||||||:||         ||:||:||
Db 66852  CAAGTGTGGGGTGTCGTATTCTAGGAGTTCTTCAGAATCCACAGCCCATATGCC---C 66796

Qy   106  LysLeuHisGlnProArgVal------------------------LeuHisCysHisVal 117
          ::|||                                       :||
Db 66795  CTGAGCAATGTCACCTCCTGAGCAT-----------------GACCCACGGTCAGCTGTGGGT 66751

Qy   118  GlnAlaArgHisGlyCysProGlyProGlnLeuLeuAlaLeuProAlaLaAspAlaAsnVal 137
          |||||||||         ||:                    ::|
Db 66750  GAGCTGGG-GCCCCACGGCGCACCCAGGTTCCTTCGTCGACTCCCACGGTGCCTTTCATG 66692

Qy   138  HisGlnLeuHisHisHisGlyCysProSerLeuArgLeuArgAlaGlnCys 157
          |||||||||         |||          ::|||||:|||
Db 66631  ATCCGTTGCACCATCCCCCACGTGCCCACCTGTTTTCCACAGTGCCCTACTGCTAAATGT 66572

Qy   158  HisArgArgProProLeuGlyLeuGlnLeuGln-LeuArgAlaPhe-------------- 172
          |||                              ||||||
Db 66571  CACTGTAGTGCACCCTGGCGGGGCGCCAACAGCCTTCTAGCCTTTGTTGGCGGAACGAAT 66512

Qy   173  --------------------HisAsnArgAlaHisHisHisGlnAl 180
                              |||:::::
Db 66511  CAGTAAGTCGTCGTCCTCGCCCCTCTCAACCTTCATTCCAAAGN---TCATTCCNT 66455

Qy   180  aArgProSerArgArgArgAlaProLysProPro 191
          :::|||         |||||||
Db 66454  CCCCCCCCANCCCAAAACACCTCACCCCCATCCTCCT 66421
```

RESULT 13
AE012121
LOCUS   AE012121   10770 bp   DNA   linear   BCT 23-MAY-2002
DEFINITION   Xanthomonas campestris pv. campestris str. ATCC 33913, section 29
             of 460 of the complete genome.
ACCESSION   AE012121 AE008922
VERSION   AE012121.1 GI:21111210
KEYWORDS   .
SOURCE   Xanthomonas campestris pv. campestris str. ATCC 33913
  ORGANISM   Xanthomonas campestris pv. campestris str. ATCC 33913
             Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
             Xanthomonas.
REFERENCE   1 (bases 1 to 10770)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
             Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
             Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
             Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
             Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
             Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
             Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
             Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
             Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
             Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
             Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
             Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
             Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
             Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
             Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
             Kitajima,J.P.
  TITLE   Comparison of the genomes of two Xanthomonas pathogens with
             differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
  MEDLINE   22022145
   PUBMED   12024217
REFERENCE   2 (bases 1 to 10770)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
             Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
             Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
             Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
             Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
             Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
             Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
             Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
             Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
             Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
             Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
             Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
             Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
             Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
             Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
             Kitajima,J.P.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de
             Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
             Brazil

FEATURES   Location/Qualifiers
     source   1..10770
                     /organism="Xanthomonas campestris pv. campestris str. ATCC
                     33913"
                     /strain="ATCC 33913"
                     /db_xref="ATCC:33913"
                     /db_xref="taxon:190485"
                     /note="pathovar: campestris"
     gene   complement(185..1795)
                     /gene="accD"
                     /note="XCC0245"
                     /note="accD"
     CDS   complement(185..1795)
                     /gene="accD"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA carboxyltransferase beta chain"
                     /protein_id="AAM39564.1"
                     /db_xref="GI:21111211"
                     /translation="MSVINSQLQVGGESFQHNAAAMRAVVEDLRATLARTALGGSEAA
                     RAKHTARGKLLVRARIDALLDPGGALLEIAPLAAHGMYDDQVPCAGVVAGIGRVSGVE
                     CVIVANDATVKGGTYYPMTVKKHLRAQEIAQONRLPCIIVLVDSGGAFLPLQDEVFDR
                     DHFGRIFYNQANLSAQGIPQIACVWGSCTAGGAYVPAMSDETVIVREQGTIFGGPPL
                     VKAATGEEVSAEELGGADVHTRISGVADHFADNDLQALARVRAIIAQLNWRKPALSMA
                     VQAPLPRHAADELGVGIPADTRRYDVREVIARIVDDSRLDEFKPRYGSTLVTGFAH
                     LHGPVGIIANGLLFSESALKGAHFIELCTQIRGIPLVPLQNITGFMVGGKYEREGIA
                     KDGAKLVMAVACAVVPKFVIGSFGACNYGMGGRAYSPNFLMWWPNARIGVMGGEQ
                     AASVLATVRRDGIEAKGGAVSGDDEEAFKSPIRAQFEQGHPYASARLWDDGIIDPA
                     DTRRVLGLGLSAALNAPIDATRFGVFRM"
     gene   complement(1806..2969)
                     /gene="accA"
                     /note="XCC0246; acd"

CDS   complement(1806..2969).
         /gene="acdA"
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="acyl-CoA dehydrogenase"
         /protein_id="AAM39565.1"
         /db_xref="GI:21111212"
         /translation="MHVPSLNFELGEDILRESVAAFASHHIAPLAAAADHDNVFPA
         QLWRLFGEQGLLGLTVEEAYGGSGMGYLAHVVAMEEISRAGCAIGLSYGAHSNICLNQ
         LRNATHEQKQRYLPKLCTGEHVGALAMASEAGSGSDVVSMKLRAEEARGDRFVLNGSKM
         WITNGPDADLVVYAKTDPSAGARGITAFIVEKGMPGFSTAQKLDKLGMRGSNTCELV
         FTDCEVPAENVLGTLNGGVRVLMSGLDPERVVLAGGPLGLMAAAMDVVLPYVHERKQF
         GEPIGTEFQLMQAKLADMYVGLNANCRAVYVYARACDAGRTRQDAAGAILYAAEKATW
         LTGQAIQVLGGNGYINEYPTGRLWRDAKLYEIGAGTSEIRRMLIGRELFERIA"
3098   3715
gene   /gene="XCC0247"
         complement(3098..3715)
3098..3715
CDS   /gene="XCC0247"
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="transcriptional regulator acrR family"
         /protein_id="AAM39566.1"
         /db_xref="GI:21111213"
         /translation="MAYRRSALMEERLAGNRERILHAARALIAEGGYRNAPVTAVAAA
         AGVSTGQIYRHFPSKADLFVEVLNAAVQHEMGILRAISAEPVSAAERLRKAIAAFVR
         ALAGPALAYAEPVEAEPVEGDEAERIRGRRLPFGEVFRQVLAEGVAAGEFPEOSLDAAAA
         CIVGAFTEALVGPVAPSRHDPQQGEOLVEAICGFCLRAAGARGPG"
3774   4175
gene   /gene="XCC0248"
         3774..4175
CDS   /gene="XCC0248"
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="conserved hypothetical protein"
         /protein_id="AAM39567.1"
         /db_xref="GI:21111214"
         /translation="MHSPCRSRAMVAVLNPLALTERLRVPDRTILSPSAMAGLLDLSQ
         QELAELAGVHRNSLRVHPESPRVQDLLRNLSRLLVAMAQIOPDEROLVFHLKNTPIPA
         FEFAPLLEVVKQGRTDDALTYLRTVAAGAAG"
4172   4708
gene   /gene="XCC0249"
         4172..4708
CDS   /gene="XCC0249"
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="conserved hypothetical protein"
         /protein_id="AAM39568.1"
         /db_xref="GI:21111215"
         /translation="MKLSAPAQCTLYRAFTPRNAAEPLSGAGAARSGGRFNRFGOPAL
         YLSLHLDTAAAEYAQAAPFPLPFFTLVSRADLPALADVRLLDASWDALWADWTDDWRK
         ALVNKVEPVSWVLGDLLREAAIPGVIFPSMAQPGGVNVALFLDMLQPEQVLHVLDDGR
         LPRDGRSWADPAAERPPE"
5006   5341
gene   /gene="XCC0250"
         complement(5006..5341)
CDS   /gene="XCC0250"
         complement(5006..5341)
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="conserved hypothetical protein"
         /protein_id="AAM39569.1"
         /db_xref="GI:21111216"
         /translation="MSVVLLALNFSGFAALCLAMEKHQHEVRGRALGMTRSRQLRAL
         GWLLLLVTPALAVSAQGWGIGSVLWLGTLSAGAAVLSLWLLPYRGGILPAAIAPAVV

gene   AGMAFALLH"
         complement(5338..6954)
         /gene="XCC0251"
CDS   complement(5338..6954)
         /gene="XCC0251"
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="conserved hypothetical protein"
         /protein_id="AAM39570.1"
         /db_xref="GI:21111217"
         /translation="MKQGFRQSMAWLHTWTGLLVGWVLLLIFMGGTASYRDELSRWM
         RPELPTTVSSAVAMRSAERVLQTIHAPDAQSWNITLPDARTPVWPAVRYMWWQNPAPPGKT
         LSRRELYGNAIIDPATGNPISARDTLGGDFFYRLHFDLHIYLPVVWARYIVGFCAMFML
         VAIISGVITHKKIFKDPFTERPGKGLRSWLDFHNVSAVTALPYHAMITYTGIVTLMFM
         YLPWGIKAQYFPDNEMRFYEESANRVADTRTFAAGTPARMRPLEEFVARARSDWRGDVG
         TVAVSLPNDAHAAVGVTQRADDLSNDAPSIILYDAVSGQRLOYSGAPGGASQTRGVNVG
         LHIAHFAGGWHRALFFGSGLLGCLMVASGVVMWAVKERPKHAKAGRIGFGLRLVDALN
         IGTVAGLPIAFAAFFWANRLVPVALEARAAMEAHLFFAAVGGALLAAFWPRRAMWSW
         QLYLGAGLFALVPLLNALTTDLHLGVTVPACQWALAGVDLVCLGLCVCLGIAGWRLQH
         WKAPQSAAARARAQAITPAPQQNALPVQEGA"
6951   7262
gene   /gene="XCC0252"
         complement(6951..7262)
CDS   /gene="XCC0252"
         complement(6951..7262)
         /note="identified by sequence similarity; putative; ORF
         located using Blastx/Glimmer/Genemark"
         /codon_start=1
         /transl_table=11
         /product="conserved hypothetical protein"
         /protein_id="AAM39571.1"
         /db_xref="GI:21111218"
         /translation="MSMASPLPAPWFKRPWLGVLGRSVAAIGGGYALASATNLLLAVV
         LPMPRSEAVLTSMLVGIVVCACAPLWAFATASVLRAWVGIGVPAVAMFAAAAMLQRGA
         A"
gene   complement(7281..7928)
         /gene="XCC0253"
CDS   complement(7281..7928)
         /gene="XCC0253"
         /note="putative; ORF located using Glimmer/Genemark"

Alignment Scores:
Pred. No.:
Score: 1.52e+03.   Length: 10770
                     149.50   Matches: 71
Percent Similarity: 37.77%   Conservative: 17
Best Local Similarity: 30.47%   Mismatches: 84
Query Match: 14.05%   Indels: 62
DB: 1   Gaps: 13

US-09-921-994-2 (1-192) x AE012121 (1-10770)

Qy   2 ThrProGlnLeuLeuLeuAlaLeuValLeuTrpAlaSerCysProAlaArgSerGlyArg   21
        |||||| ||| ||| ||||||||||||
Db 5023 ACGCCCATCCGGCGACGACGAGGTGCAGCAAATGG-------CTGCCGGCAAGATCAGGCGG   5076

Qy   22 LysGlyProProAla------AlaLeuThrLeuProArgValGlnCysProAlaLeuSerArg   39
        ||| || ||||||||||::::: |||
Db 5077 GCCGGTACGGCGCAGGAACCACAGCGACACACCGCGGCCGCCGCCGCG-------TCACG   5127

Qy   40 TyrProIleAlaValAspCysSerTrpThrLeuProHisAspAlaAlaAlaSerProGly   59
        ||| ||| ||| |||| ::||
Db 5128 TGCCCAGCCACAGTACGCTGC----------------------CGATGCCCAGCCTTCGCGCCACA   5172

Qy   60 ProCysProLeuGlyGlnLeuProAlaLeuArgTrpLysGluArgAla-----ProSer   77
        ||| ||||| |||||||||:::: |||:::
Db 5173 CCG--------------CCAGCCGCAAAGGTGACCAGCAACACAGCGACCATCCCAGCG   5214

Qy   78 SerSerAspThrAlaProGlyAlaMetProSerLeuSerValProAlaspArgArgGlyLeu   97
        ::: |||||||| |||:::: ||| |||:::
Db 5215 CGCCGCAACTGACCTGAGCGGGTGCATGCGCCCCCCCACGCACCTCGTTCCTGGTGTT   5274

Qy   98 LeuLeuAspProAla--   102

```
Db  5275 TTTCCATCCGCCAGGCAACAAGGCCGCGGCGAAGCCGGGAAAAATTCAAGGCAGCCAGCAACAGCACTA 5334
              |||
Qy  103  ---AlaCysSerLysLeuValHisGlnProArgValLeuHisCysHisValGlnAlaArgHis 121
                 |:::    :|||:::    ||||||     |||
Db  5335 CGCTCAATGCCGCCTTCCTGCACAGGACGACGGCGGTTTTGCGGCGCAGGCGTGGCTTGCGGCGCGG 5394
                                              |||
Qy  122  Gly-------------------CysProGly----------- 134
                  ||| ||| ||| ||| ||| ||| |||
Db  5395 GGCCTTGCGTGCCGCGGGAGTTGCGGCGCCTTCCAGTGTGCTGCCAGGGCCCGCCGGAT 5454
                                                 |||
Qy  135  --------------AlaAsnValHisGlnLeuHisHisHisGlyCysPro---Ala 147
                    ||| |||||||||:::||    |||
Db  5455 GCCCAGGCGACACCCCAGGCACACCCAGGCGCGCCCAGGCCCAATGCCCATTGGCC 5514
                                    |||||||     |||
Qy  148  ValLeuHisGlySerLeuArgAlaGlnCys----------------HisArg 159
              |||         |:::||
Db  5515 TGCCGGCGACGGTGCACGGGACGTGCACGGCGTCGGTGGGTCAGCGCGGCAGCACCA 5574
                                                  |||
Qy  160  ArgProProLeuGlyLeuGlnGlnGlnLeuLeuArgAlaPheHisAsnArgAlaHisHisGln 179
                                                         |||
Db  5575 CGCGAACAAGCCCTGCGCCCAGGTTACAGCTGCCCAGGCTCCACATCG--------CACGCC 5624
                   :::        |||       |||
Qy  180  AlaArgProSerArgArgArgAlaProLysProProArg 192
              ||||||||||     :::    |||
Db  5625 GCGGCCACACGACGAAGGCGCGAGCAAGGCGGCCGGCACCCCAG 5663

RESULT 14
LOCUS       AC131212               168964 bp    DNA    linear    HTG 18-AUG-2002
DEFINITION  Homo sapiens clone RP13-554M15, *** SEQUENCING IN PROGRESS ***, 23
            unordered pieces.
ACCESSION   AC131212
VERSION     AC131212.1  GI:22296917
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168964)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C.C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Deferich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,S., Oguh,M., Okwuonu,G.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Nguyen,A., Nguyen,N.,
            Oraguzie,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,E., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Role,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
```

```
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168964)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     --------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
            Center project name: HEME
            Center clone name: RP13-554M15
            --------------- Summary Statistics
            Sequencing vector: plasmid;
            Chemistry: Dye-terminator Big Dye; version 0.990329
            Assembly program: Phrap; version 149776 bases at least Q40
            Consensus quality: 152075 bases at least Q30
            Consensus quality: 153483 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1:  contig of 2409 bp in length
            *       2410:  gap of unknown length
            *       2409:  contig of 2409 bp in length
            *       2410:  gap of unknown length
            *       2509:  contig of 2383 bp in length
            *       2510:  gap of unknown length
            *       4892:  contig of 2202 bp in length
            *       4893:  gap of unknown length
            *       4993:  contig of 2202 bp in length
            *       7194:  gap of unknown length
            *       7195:  contig of 2153 bp in length
            *       7294:  gap of unknown length
            *       7295:  contig of 2153 bp in length
            *       9447:  gap of unknown length
            *       9448:  contig of 3363 bp in length
            *       9547:  gap of unknown length
            *       9548:  contig of 3363 bp in length
            *      12910:  gap of unknown length
            *      12911:  contig of 3186 bp in length
            *      13010:  gap of unknown length
            *      13011:  contig of 3186 bp in length
            *      16196:  gap of unknown length
            *      16197:  contig of 3263 bp in length
            *      16296:  gap of unknown length
            *      16297:  contig of 3263 bp in length
            *      19559:  gap of unknown length
            *      19560:  contig of 4352 bp in length
            *      19659:  gap of unknown length
            *      19660:  contig of 4352 bp in length
            *      24011:  gap of unknown length
            *      24012:  contig of 4658 bp in length
            *      24111:  gap of unknown length
            *      24112:  contig of 4658 bp in length
            *      28769:  gap of unknown length
            *      28770:  contig of 5682 bp in length
            *      28869:  gap of unknown length
            *      28870:  contig of 5682 bp in length
            *      34551:  gap of unknown length
            *      34552:  contig of 4564 bp in length
            *      34651:  gap of unknown length
            *      34652:  contig of 4564 bp in length
            *      39215:  gap of unknown length
            *      39216:  contig of 4353 bp in length
            *      39315:  gap of unknown length
            *      39316:  contig of 4353 bp in length
            *      43668:  gap of unknown length
            *      43669:  contig of 6999 bp in length
            *      43768:  gap of unknown length
            *      43769:  contig of 6999 bp in length
            *      50767:  gap of unknown length
            *      50768:  contig of 6985 bp in length
            *      50867:  gap of unknown length
            *      50868:  contig of 6985 bp in length
            *      57852:  gap of unknown length
            *      57853:  contig of 6546 bp in length
            *      57952:  gap of unknown length
            *      57953:  contig of 6546 bp in length
            *      64498:  gap of unknown length
            *      64499:  contig of 6590 bp in length
            *      64598:  gap of unknown length
            *      64599:  contig of 6590 bp in length
            *      71188:  gap of unknown length
            *      71189:  contig of 6590 bp in length
            *      71288:  gap of unknown length
```

```
LOCUS       AC079031     186558 bp    DNA    linear   HTG 31-JUL-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-503G7, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC079031
VERSION     AC079031.23  GI:22023367
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186558)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C.C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Deford,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,E., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojukokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,I.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G., and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186558)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 186558)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 29, 2002 this sequence version replaced gi:21908302.
            -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Drafting Center Code: BCM
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
```

```
FEATURES             Location/Qualifiers
     source          1..168964
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP13-554M15"

Homo sapiens
     1  (bases 1 to 168558)

        71289   80962:  gap of 9674 bp in length
        80963   81062:  contig of unknown length
        81063   92067:  gap of 11006 bp in length
        92068   92167:  contig of unknown length
        92168  105140:  gap of 12972 bp in length
       105141  105240:  contig of unknown length
       105241  119456:  gap of 14216 bp in length
       119457  119556:  contig of unknown length
       119557  133120:  gap of 13564 bp in length
       133121  133220:  contig of unknown length
       133221  150729:  gap of 17509 bp in length
       150730  150829:  contig of unknown length
       150830  168964:  gap of 18135 bp in length.

BASE COUNT    36032 a  46964 c  48094 g  35152 t  2722 others
ORIGIN
```

```
US-09-921-994-2 (1-192) x AC131212 (1-168964)

Alignment Scores:
Pred. No.:
Score:                 1.69e+04          Length:        168964
                       149.00            Matches:       66
Percent Similarity:    39.59%            Conservative:  12
Best Local Similarity: 33.50%            Mismatches:    77
Query Match:           14.00%            Indels:        42
DB:                    2                 Gaps:          10
```

```
Qy   12  TrpAlaSerCysPro----ProArgSerGlyArgGlyLysGlyPro-ProAlaAlaLeuThrIle  30
             ||  |||                  ||||||||| |||||||  |||||
Db 127607  TGGTGCGGCACCCGTCCCTCTGCCTCCCGTGCCCCAGCGCGGCACACTT----           127661

Qy   30  uProArgValGlnCysArgAlaSerArgTyrProIleAlaValAspCysSer-TrpThrI  50
               :::|| |||||||  :|||||||:::|||
Db 127662  -----GCCGAGGAGAGGGCCTTGCGGCACCCGGCACCCAGCACTGCCCTGTGG----       127710

Qy   50  euProHisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuA  70
            |||||  |||||  |  |||  :::
Db 127711  --CCTCCGTGACCTGCTTACGGTAAAAGCCCAGTACCCAGAGGAGGCGGCCGCTCA        127768

Qy   70  rgTrpLysGlu----ArgAlaProSerSerSerAspThrAlaProGlyAlaMetPros  88
            ||||||||  |||  :::
Db 127769  CGCGTGAGGATCACACCGTGTGACACACAGCCGCTCATGCCTGGCCT-                127827

Qy   88  erLeuSerValProAspArgArgGlyLeuLeuLeuAspProAlaAlaCysSerSerLysLeuH  108
            |||            |||  ::::
Db 127828  ----TCAGGGCCCGACCAGCGCTGGAGTC----                              127870

Qy  108  lsGlnProArgValLeuHis------  115
             |||
Db 127871  CCCAGCGCCAGGTGGTCTCACCTTTATTCTTCTCGTGACGGTAGATCTCCCAGCGCGATC--C 127930

Qy  115  ysHisValGlnAlaArgHisSGlyCysProGlyProGlnLeuAlaLeuProAlaAlaAspA  135
             ||||||
Db 127931  TGCACGTG-----CACGGCACACCACCTCCCCGGCTTCTGCGACAGGATGGAAGCG        127981

Qy  135  laAsnValHisGlnLeuHisHisGlyCys-----ProAlaValLeuHisG  151
             ||                             ||||||||||
Db 127982  CTCAGTGGGGCGGCTGGTGTCCACCCCACCCATGTTTTCTCGAGACTCCTGCTGTCACGGGCC 128041

Qy  151  lysSerLeuArgAlaGlnCysHisSerArgArgProLeuGlyLeuGlnLeuGlnLeuArgA  171
             :::              |||  ::::||||||||
Db 128042  CCCCGGAGGCGGCCCGCCGGGGGGCAGAGTCCTCCCTTGCCA                      128086

Qy  171  laPheHisAsnArgAlaHisHisGlnAlaArgProSerArgArgArg  186
             |||||||||||||                         ||| |||
Db 128087  CCTTCCAACACGTGCTCACCTTGACCACCACGCCCGGTACGGGTCAGA                128133
```

```
RESULT 15
AC079031
```

Center project name: HBQO
Center clone name: RP11-503G7
---------- Summary Statistics

Sequencing vector: M13;
Sequencing vector: Plasmid;
Chemistry: Dye-primer BigDye: 19% of reads
Chemistry: Dye-terminator Big Dye: 81% of reads
Assembly program: Phrap: version 0.990329
Assembly program: Phrap; version 0.990329
Consensus quality: 185354 bases at least Q40
Consensus quality: 185593 bases at least Q30
Consensus quality: 185690 bases at least Q20
Consensus quality: 185621; sum-of-contigs estimation
Estimated insert size: 185621; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

----

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/GenBank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1   61396: contig of 61396 bp in length
*   61397   61496: gap of unknown length
*   61497   186558: contig of 125062 bp in length.

FEATURES   Location/Qualifiers
source   1..186558
                  /organism="Homo sapiens".
                  /db_xref="taxon:9606"
                  /chromosome="12"
                  /clone="RP11-503G7"

BASE COUNT   38738 a   55666 c   57117 g   34936 t   101 others
ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.84e+04   Length: 186558
Percent Similarity: 149.00   Matches: 66
Best Local Similarity: 39.59%   Conservative: 12
Query Match: 33.50%   Mismatches: 77
                       14.00%   Indels: 42
DB: 2   Gaps: 10

US-09-921-994-2 (1-192)   x   AC079031 (1-186558)

Qy   12 TrpAlaSerCysPro---ProArgSerGlyArgLysGlyPro-ProAlaAlaLeuThrLe 30
         |||:   ::||||||||| ||||||
Db 101596 TGGTGCGGCACCCCGTGCCCCTGCTCGCTCCCCGGCCCCGGCCACACTT----- 101650

Qy   30 uProArgValGlnCysArgAlaSerArgTyrProIleAlaValAlaAspCysSer-TrpThrL 50
            ::|| |||||| |||||| |||||| |||
Db 101651 -------GCCGAGGAGGGCCCTTGCCGGCCACCCGGCCACTGCCCTGTGG---- 101699

Qy   50 euProHisAspProAlaAlaSerProGlyProCysProLeuGlyGlnLeuProAlaLeuA 70
            |||||| |||||| ||||||| ||||:::
Db 101700 --CCTCGCCTGACCTGCTTACGGTAAAAGCCCAGTACCCACAGAGGAAGGGCGAGGGGCTCA 101757

Qy   70 rgTrpLysGlu-----ArgAlaProSerSerSerAspThrAlaProGlyAlaMetPros 88
         :::||:   :::
Db 101758 CGCGTGGAGGAATCACACCCGTGCACACGCGTGTGAACACACACCCTCATGCCTGGCCT- 101816

Qy   88 erLeuSerValProAspArgArgGlyLeuLeuLeuAspProAlaAlaCysSerLysLeuH 108
         ||||||: |||||||::: ||||:::
Db 101817 ----TCAGGGCCCGACCAGCCTGGAGTC----GGGTTCACAAGCCCCTGC 101859

Qy   108 isGlnProArgValLeuHis----------------- 115
         ||||||||||||
Db 101860 CCCAGCCCCAGGTGCTCACCTTTATCTTCTGCTGGTGACGGTAGATCTGCCAGCCGATC 101919

Qy   115 ysHisValGlnAlaArgHisGlyCysProGlyProGlnLeuAlaLeuProAlaAlaAspA 135
         |||||| ||| |||||||::||
Db 101920 TGCACGTG------------CACGGCACACCACCTCCCCGGCTTCTGCGACAGGATGGAAGCG 101970

Qy   135 laAsnValHisGlnLeuHisHisHisGlyCys-------------------ProAlaValLeuHisG 151
         ||| ||||||| ||||| ||||||||||
Db 101971 CTCAGTGGGGCTGGTGCCACCCCACCCCACCCCATGTTTTTCTGCAGACTCCTTGCTCGTCACGGGCC 102030

Qy   151 lySerLeuArgAlaGlnCysHisArgArgProProLeuGlyLeuGlnLeuGlnLeuArgA 171
            ::: ||||||||||||||||||:
Db 102031 CCCCGAGGCCGCGCCCCCGGGGCGACCAGAGTCCTTCCCTTGCCA-------------CGGG 102075

Qy   171 laPheHisAsnArgAlaHisHisGlnLeuAlaArgProSerArgArgArgArgA 186
         |||| |||||||| ||| |||
Db 102076 CCTTCCCAACACGTTGCTCACCTTGACCACCCGCCCGGTTACGGGGTCAGA 102122

Search completed: December 28, 2002, 12:30:54
Job time : 1865.5 secs